Journal of Experimental Psychology

David A. Grant, Editor
Department of Psychology, University of Wisconsin
Madison, Wisconsin 53706

James E. Deese, Associate Editor
Johns Hopkins University

Harold W. Hake, Associate Editor
University of Illinois, Urbana

Delos D. Wickens, Associate Editor
Ohio State University

Consulting Editors

Abram Amsel, *University of Toronto*
Fred Attneave, *University of Oregon*
William F. Battig, *University of Colorado*
Lyle E. Bourne, Jr., *University of Colorado*
Gordon H. Bower, *Stanford University*
George E. Briggs, *Ohio State University*
Cletus J. Burke, *California State College, Hayward*
Virgil R. Cartson, *National Institute of Mental Health*
Tom N. Cornsweet, *University of California, Berkeley*
John W. Cotton, *University of California, Santa Barbara*
Kent Dallett, *University of California, Los Angeles*
M. R. D'Amato, *Rutgers University*
Sheldon M. Ebenholtz, *University of Wisconsin*
Ward Edwards, *University of Michigan*
Charles W. Eriksen, *University of Illinois, Urbana*
Raymond W. Frankmann, *University of Houston*
Julian Hochberg, *New York University*
Leonard M. Horowitz, *Stanford University*
William C. Howell, *Ohio State University*
Stewart H. Hulse, *Johns Hopkins University*
Arthur L. Irion, *Tulane University*
Geoffrey Keppel, *University of California, Berkeley*
Gregory A. Kimble, *Duke University*
Herbert D. Kimmel, *Ohio University*
Edmund Klemmer, *Bell Telephone Laboratories, Holmdel, New Jersey*
Harlan L. Lane, *University of Michigan*
Herschel W. Leibowitz, *Pennsylvania State University*
Donald B. Lindsley, *University of California, Los Angeles*
Frank A. Logan, *University of New Mexico*
Irving Maltzman, *University of California, Los Angeles*
Arnold Mechanic, *California State College, Hayward*
Bennet B. Murdock, Jr., *University of Toronto*
Merrill E. Noble, *Kansas State University*
Michael I. Posner, *University of Oregon*
Leo Postman, *University of California, Berkeley*
L. Starling Reid, *University of Virginia*
Donald A. Riley, *University of California, Berkeley*
Leonard E. Ross, *University of Wisconsin*
Rudolph Schulz, *State University of Iowa*
Harold W. Stevenson, *University of Minnesota*
David R. Thomas, *University of Colorado*
Milton A. Trapold, *University of Minnesota*
Endel Tulving, *University of Toronto*

Helen Orr
Managing Editor

Evelyn Kilsheimer
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Robert D. Gorman
Subscription Manager

Volume 74, 1967

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 74

Agee, F. L., Jr., and Gogel, W. C. Equidistant Judgments in the Vicinity of a Binocular Illusion 87
Agnew, H. J., Jr. See Webb, W. B.
Amsel, A. See Donin, J. A.
Arbuckle, T. Y. Differential Retention of Individual Paired Associates within an R T "Learning" Trial 443
Attneave, F., and Olson, R. K. Discriminability of Stimuli Varying in Physical and Retinal Orientation 149
Bartz, W. H., Satz, P., and Fennell, E. Grouping Strategies in Dichotic Listening: The Effects of Instructions, Rate, and Ear Asymmetry 132
Bem, S. L. Verbal Self-Control: The Establishment of Effective Self-Instruction 485
Bernstein, I. H., Schurman, D. L., and Forester, G. Choice Reaction Time as a Function of Stimulus Uncertainty, Response Uncertainty, and Behavioral Hypotheses .. 517
Bevan, W. See Duke, W. F.
Bloch, R., and Voss, J. F. Acquisition of Paired Associates as a Function of Associative Strength of Competing Responses 14
Borresen, C. R. See Lichte, W. H.
Boudewyns, P. A. See Madison, H. L.
Bridger, W. H. See Mandel, I. J.
Brown, S. C., and Rubin, E. D. Cue Utilization in Serial Learning 550
Bryson, M. R. See Feather, B. W.
Burstein, K. R., Epstein, S., and Smith, B. Primary Stimulus Generalization of the GSR as a Function of Objective and Subjective Definition of the Stimulus Dimension 124
Camp, D. S., Raymond, G. A., and Church, R. M. Temporal Relationship between Response and Punishment 114
Capaldi, E. J. Sequential versus Nonsequential Variables in Partial Delay of Reward .. 161
Carlton, R. A. See Prokasy, W. F.
Church, R. M. See Camp, D. S.
Cobb, N. J. See Goss, A. E.
Collins, J. F. See Eriksen, C. W.
Corballis, M. C. Serial Order in Recognition and Recall 99
Corman, C. D. Stimulus Generalization of Habituation of the Galvanic Skin Response 236
Crowder, R. G. Proactive and Retroactive Inhibition in the Retention of a T-Maze Habit in Rats 167
Danks, J. H. See Glucksberg, S.
Davidov, W. See Kimmel, H. D.
Delse, F. C. See Feather, B. W.
Desiderato, O., and Wassarman, M. E. Incubation of Anxiety: Effect on Generalization Gradients 506
Devine, J. V. See Haygood, R. C.
Dominowski, R. L., and Ekstrand, B. R. Direct and Associative Priming in Anagram Solving 84
Donin, J. A., Surridge, C. T., and Amsel, A. Extinction Following Partial Delay of Reward Interpolated, at 24-Hour Intertrial Intervals 50
Dufort, R. H. Eyelid Conditioning as a Function of UCS Duration with Drive Equated 321
Duke, W. F., and Bevan, W. Stimulus Variation and Repetition in the Acquisition of Naming Responses 178
Egeth, H., and Smith, E. E. Perceptual Selectivity in a Visual Recognition Task 543
Eisenberger, R., Karpman, M., and Trattner, J. What is the Necessary and Sufficient Condition for Reinforcement in the Contingency Situation? 342
Ekstrand, B. R. See Dominowski, R. L.
Ekstrand, B. R. See Underwood, B. J.
Epstein, S., and Smith, B. Modes and Adequacy of Resolution of Three Basic Types of Cognitive-Motor Conflict 264
Epstein, S. See Burstein, K. R.
Erickson, J. R., and Zajkowski, M. M. Learning Several Concept-Identification Problems Concurrently: A Test of the Sampling-with-Replacement Assumption 212
Eriksen, C. W., and Collins, J. F. Some Temporal Characteristics of Visual Pattern Perception 476
Feather, B. W., Delse, F. C., and Bryson, M. R. Human Salivary Conditioning: Effect of Unconditioned-Stimulus Intensity 389

Fennell, E. See Bartz, W. H.
Forester, G. See Bernstein, I. H.
Forrester, W. E., and Spear, N. E. Coding Processes in Verbal Learning as a Function of Response Pronounciability ... 586
Friedman, M. J., and Reynolds, J. H. Retroactive Inhibition as a Function of Response-Class Similarity ... 351
Gescheider, G. A., and Niblette, R. K. Cross-Modality Masking for Touch and Hearing 313
Glucksberg, S., and Danks, J. H. Functional Fixedness: Stimulus Equivalence Mediated by Semantic-Acoustic Similarity ... 400
Goff, G. D. Differential Discrimination of Frequency of Cutaneous Mechanical Vibration ... 294
Gogel, W. C. See Agee, F. J., Jr.
Goss, A. E., and Cobb, N. J. Formation, Maintenance, Generalization, and Retention of Response Hierarchies: The Role of Meaningfulness of Response Members ... 272
Gould, J. D., and Schaffer, A. Eye-Movement Parameters in Pattern Recognition ... 225
Greeno, J. G. See James, C. T.
Gregg, L. W. See McLean, R. S.
Grice, G. R., Hunter, J. J., Kohfeld, D. L., and Masters, L. Order of Presentation, CS Intensity, and Response Latency ... 581
Grote, F. W. See Strickland, L. H.
Halpern, J., and Moore, J. W. Two-Choice Discrimination Learning as a Function of Cue Similarity and Probability of Reinforcement ... 182
Harbin, S. P., and Wright, J. H. Connotative Meaning in Verbal Learning ... 106
Hardacre, L. E. See Homzie, M. J.
Harding, G. B. See Wickens, D. D.
Harvey, L. O., Jr. See Leibowitz, H. W.
Hastings, S. E., and Obrist, P. A. Heart Rate during Conditioning in Humans: Effect of Varying the Interstimulus (CS-UCS) Interval ... 431
Haygood, R. C., and Devine, J. V. Effects of Composition of the Positive Category on Concept Learning ... 230
Haygood, R. C., and Stevenson, M. Effects of Number of Irrelevant Dimensions in Nonconjunctive Concept Learning ... 302
Hersen, M. Experimentally Induced Response Biases as a Function of Positive and Negative Wording ... 588
Higgins, J. D. See Prokasy, W. F.
Homzie, M. J., Hardacre, L. E., and Weimer, J. Transfer of Training as a Function of Stimulus and Response Integration of Paired Associates ... 383
Houston, B. K., and Jones, T. M. Distraction and Stroop Color-Word Performance ... 54
Houston, J. P. Unlearning of Specific Associations in the A-B, C-D Paradigm ... 254
Hunt, D. P. Differential GSR Conditioning at Extended Interstimulus Intervals ... 20
Hunter, J. J. See Grice, G. R.
Ison, J. R. See Pennes, E. S.
James, C. T., and Greeno, J. G. Stimulus Selection at Different Stages of Paired-Associate Learning ... 75
Johnston, W. A. S-R, R-S Independence and the Interference Potency of Latent R-S Associations ... 511
Jones, T. M. See Houston, B. K.
Karpman, M. See Eisenberger, R.
Karsh, R. See Monty, R. A.
Kasschau, R. A., and Pollio, H. R. Response Transfer Mediated by Meaningfully Similar and Associated Stimuli Using a Separate-Lists Design ... 146
Keppel, G., Zavortink, B., and Schiff, B. B. Unlearning in the A-B, A-C Paradigm as a Function of Percentage Occurrence of Response Members ... 172
Keppel, G. See Postman, L.
Kimmel, H. D., and Davidov, W. Classical GSR Conditioning with Concomitant EMG Measurement ... 67
Kimmel, H. D., and Sternthal, H. S. Replication of GSR Avoidance Conditioning with Concomitant EMG Measurement and Subjects Matched in Responsivity and Conditionability ... 144
Kohfeld, D. L. See Grice, G. R.
Krudis, B. R. See Nakamura, C. Y.
Lehr, P. J. See Mikulka, P. J.
Lehr, R. See Pavlik, W. B.
Leibowitz, H. W., and Harvey, L. O., Jr. Size Matching as a Function of Instructions in a Naturalistic Environment ... 378
Leibowitz, H. W. See Meneghini, K. A.

LEITENBERG, H. **Punishment Training with and without an Escape Contingency** 393
LICHTE, W. H., AND BORRESEN, C. R. **Influence of Instructions on Degree of Shape Constancy** 538
LOESS, H. **Short-Term Memory, Word Class, and Sequence of Items** 556
MADISON, H. L., AND BOUDEWYNS, P. A. **Asymmetric-Reinforcing Events in Probability Learning** 492
MAIER, S. F. See SELIGMAN, M. E. P.
MANDEL, I. J., AND BRIDGER, W. H. **Interaction between Instructions and ISI in Conditioning and Extinction of the GSR** 36
MANDLER, G. See SEGAL, M. A.
MARSTON, A. **Self-Reinforcement and External Reinforcement in Visual-Motor Learning** 93
MARTIN, E. **Relation Between Stimulus Recognition and Paired-Associate Learning** ... 500
MASTERS, L. See GRICE, G. R.
MAYHEW, A. J. **Interlist Changes in Subjective Organization during Free-Recall Learning** 425
McLEAN, R. S., AND GREGG, L. W. **Effects of Induced Chunking on Temporal Aspects of Serial Recitation** 455
MENEGHINI, K. A., AND LEIBOWITZ, H. W. **Effect of Stimulus Distance and Age on Shape Constancy** 241
MIKULKA, P. J., LEHR, R., AND PAVLIK, W. B. **Effect of Reinforcement Schedules on Reward Shifts** 57
MONTY, R. A., KARSH, R., AND TAUB, H. A. **Pacing of Rehearsal in Sequential Short-Term Memory** 300
MOORE, J. W. See HALPERN, J.
MURDOCK, B. B., JR., AND VOM SAAL, W. **Transpositions in Short-Term Memory** 137
NAKAMURA, C. Y., AND KRUDIS, B. R. **Evaluation of a Response Speed Measure of Incentive Value of Reward** 44
NEIMARK, E. D. **Effect of Differential Reinforcement upon Information-Gathering Strategies in Diagnostic Problem Solving** 406
NESMITH, R., AND RODWAN, A. S. **Effect of Duration of Viewing on Form and Size Judgments** 26
NIBLETTE, R. K. See GESCHEIDER, G. A.
NOBLE, M. See SWINK, J.
OBRIST, P. A. See HASTINGS, S. E.
OLSON, R. K. See ATTNEAVE, F.
OVER, R. **Effect of the Angle of Tilt of the Inspection Figure on the Magnitude of a Kinesthetic Aftereffect** 249
PAVLIK, W. B., AND LEHR, R. **Strength of Alternative Responses and Subsequent Choices** 562
PAVLIK, W. B. See MIKULKA, P. J.
PENNES, E. S., AND ISON, J. R. **Effects of Partial Reinforcement on Discrimination Learning and Subsequent Reversal or Extinction** 219
PLATT, J. R. See SPENCE, K. W.
POLLIO, H. R. See KASSCHAU, R. A.
POSTMAN, L., AND KEPPEL, G. **Retroactive Inhibition in Free Recall** 203
PROKASY, W. F., CARLTON, R. A., AND HIGGINS, J. D. **Effects of Nonrandom Intermittent Reinforcement Schedules in Human-Eyelid Conditioning** 282
RAYMOND, G. A. See CAMP, D. S.
REYNOLDS, J. H. See FRIEDMAN, M. J.
RODWAN, A. S. See NESMITH, R.
ROSKIND, W. See SIGEL, I. E.
ROYER, F. L. **Sequential Complexity and Motor Response Rates** 199
RUBIN, E. D. See BROWN, S. C.
SALTZ, E. See SIGEL, I. E.
SATZ, P. See BARTZ, W. H.
SAUFLEY, W. H., JR. **An Analysis of Cues in Serial Learning** 414
SCHAFFER, A. See GOULD, J. D.
SCHIFF, B. B. See KEPPEL, G.
SCHURMAN, D. L. See BERNSTEIN, I. H.
SEGAL, M. A., AND MANDLER, G. **Directionality and Organizational Processes in Paired-Associate Learning** 305
SEGAL, S. J. **Priming of Association Test Responses by Differential Verbal Contexts** ... 370
SELIGMAN, M. E. P., AND MAIER, S. F. **Failure to Escape Traumatic Shock** 1
SIGEL, I. E., SALTZ, E., AND ROSKIND, W. **Variables Determining Concept Conservation in Children** 471
SLAMECKA, N. J. **Recall and Recognition in List-Discrimination Tasks as a Function of the Number of Alternatives** 187

Slamecka, N. J. Serial Learning and Order Information 62

Smith, B. See Burstein, K. R.

Smith, B. See Epstein, S.

Smith, E. E. Effects of Familiarity on Stimulus Recognition and Categorization 324

Smith, E. E. See Egeth, H.

Spear, N. E., and Spitzner, J. H. Effect of Initial Nonrewarded Trials: Factors Responsible for Increased Resistance to Extinction 525

Spear, N. E. See Forrester, W. E.

Spence, K. W., and Platt, J. R. Effects of Partial Reinforcement on Acquisition and Extinction of the Conditioned Eyeblink in a Masking Situation 259

Sperling, S. E. Position Responding and Latency of Choice in Simultaneous Discrimination Learning 333

Spitzner, J. H. See Spear, N. E.

Sternthal, H. S. See Kimmel, H. D.

Stevenson, M. See Haygood, R. C.

Strickland, L. H., and Grote, F. W. Temporal Presentation of Winning Symbols and Slot-Machine Playing 10

Suboski, M. D. UCS Intensity and the Latency of the Classically Conditioned Eyelid Response 31

Surridge, C. T. See Donin, J. A.

Swink, J., Trumbo, D., and Noble, M. On the Length-Difficulty Relation in Skill Performance 356

Taub, H. A. See Monty, R. A.

Tighe, T. J., and Tighe, L. S. Discrimination Shift Performance of Children as a Function of Age and Shift Procedure 466

Tighe, L. S. See Tighe, T. J.

Trattner, J. See Eisenberger, R.

Trumbo, D. See Swink, J.

Turvey, M. T. Repetition and the Preperceptual Information Store 289

Underwood, B. J., and Ekstrand, B. R. Studies of Distributed Practice: XXIV. Differentiation and Proactive Inhibition 574

Underwood, B. J., and Ekstrand, B. R. Word Frequency and Accumulative Proactive Inhibition 193

vom Saal, W. See Murdock, B. B., Jr.

Voss, J. F. See Bloch, R.

Wassarman, M. E. See Desiderato, O.

Webb, W. B., and Agnew, H. J., Jr. Sleep Cycling within Twenty-Four Hour Periods .. 158

Weimer, J. See Homzie, M. J.

Wickens, D. D., and Harding, G. B. Effects of Interstimulus Interval, Forewarning, and Instructions on Extinction of the GSR 363

Willers, E. P. Nonstimulus and Nonretinal Mechanisms in Figural Aftereffects 452

Williams, R. L. Response Strength as a Function of Pre- and Postreward Delay and Physical Confinement 420

Willis, M. P. Stress Effects on Skill 460

Wright, J. H. See Harbin, S. P.

Zajkowski, M. M. See Erickson, J. R.

Zavortink, B. See Keppel, G.

Journal of Experimental Psychology

Vol. 74, No. 1 May 1967

FAILURE TO ESCAPE TRAUMATIC SHOCK[1]

MARTIN E. P. SELIGMAN[2] AND STEVEN F. MAIER[3]

University of Pennsylvania

Dogs which had 1st learned to panel press in a harness in order to escape shock subsequently showed normal acquisition of escape/avoidance behavior in a shuttle box. In contrast, yoked, inescapable shock in the harness produced profound interference with subsequent escape responding in the shuttle box. Initial experience with escape in the shuttle box led to enhanced panel pressing during inescapable shock in the harness and prevented interference with later responding in the shuttle box. Inescapable shock in the harness and failure to escape in the shuttle box produced interference with escape responding after a 7-day rest. These results were interpreted as supporting a learned "helplessness" explanation of interference with escape responding: Ss failed to escape shock in the shuttle box following inescapable shock in the harness because they had learned that shock termination was independent of responding.

Overmier and Seligman (1967) have shown that the prior exposure of dogs to inescapable shock in a Pavlovian harness reliably results in interference with subsequent escape/avoidance learning in a shuttle box. Typically, these dogs do not even escape from shock in the shuttle box. They initially show normal reactivity to shock, but after a few trials, they passively "accept" shock and fail to make escape movements. Moreover, if an escape or avoidance response does occur, it does not reliably predict future escapes or avoidances, as it does in normal dogs.

This pattern of effects is probably not the result of incompatible skeletal responses reinforced during the inescapable shocks, because it can be shown even when the inescapable shocks are delivered while the dogs are paralyzed by curare. This behavior is also probably not the result of adaptation to shock, because it occurs even when escape/avoidance shocks are intensified. However, the fact that interference does not occur if 48 hr. elapse between exposure to inescapable shock in the harness and escape/avoidance training, suggests that the phenomenon may be partially dependent upon some other temporary process.

---

[1] This research was supported by grants to R. L. Solomon from the National Science Foundation (GB-2428) and the National Institute of Mental Health (MH-04202). The authors are grateful to R. L. Solomon, J. Aronfreed, J. Geer, H. Gleitman, F. Irwin, D. Williams, and J. Wishner for their advice in the conduct and reporting of these experiments. The authors also thank J. Bruce Overmier with whom Exp. I was begun.

[2] National Science Foundation predoctoral fellow.

[3] National Institute of Mental Health predoctoral fellow.

Overmier and Seligman (1967) suggested that the degree of control over shock allowed to the animal in the harness may be an important determinant of this interference effect. According to this hypothesis, if shock is terminated independently of *S*'s responses during its initial experience with shock, interference with subsequent escape/avoidance responding should occur. If, however, *S*'s responses terminate shock during its initial experience with shock, normal escape/avoidance responding should subsequently occur. Experiment I investigates the effects of escapable as compared with inescapable shock on subsequent escape/avoidance responding.

EXPERIMENT I

Method

Subjects.—The *S*s were 30 experimentally naive, mongrel dogs, 15–19 in. high at the shoulder, and weighing between 25 and 29 lb. They were maintained on ad lib food and water in individual cages. Three dogs were discarded from the Escape group, two because they failed to learn to escape shock in the harness (see procedure), and one because of a procedural error. Three dogs were discarded from the "Yoked" control group, two because they were too small at the neck to be adequately restrained in the harness; the third died during treatment. This left 24 *S*s, eight in each group.

Apparatus.—The apparatus was the same as that described in Overmier and Seligman (1967). It consisted of two distinctively different units, one for escapable/inescapable shock sessions and the other for escape/avoidance training. The unit in which *S*s were exposed to escapable/inescapable shock consisted of a rubberized, cloth hammock located inside a shielded, white, sound-attenuating cubicle. The hammock was constructed so that *S*'s legs hung down below its body through four holes. The *S*'s legs were secured in this position, and *S* was strapped into the hammock. In addition, *S*'s head was held in position by panels placed on either side and a yoke between the panels across *S*'s neck. The *S* could press the panels with its head. For the Escape group pressing the panels terminated shock, while for the "Yoked" control group, panel presses did not effect the preprogrammed shock. The shock source for this unit consisted of 500 v. ac transformer and a parallel voltage divider, with the current applied through a fixed resistance of 20,000 ohms. The shock was applied to *S* through brass plate electrodes coated with commercial electrode paste and taped to the footpads of *S*'s hind feet. The shock intensity was 6.0 ma. Shock presentations were controlled by automatic relay circuitry located outside the cubicle.

Escape/avoidance training was conducted in a two-way shuttle box with two black compartments separated by an adjustable barrier (described in Solomon & Wynne, 1953). The barrier height was adjusted to *S*'s shoulder height. Each shuttle-box compartment was illuminated by two 50-w. and one 7½-w. lamps. The CS consisted of turning off the four 50-w. lamps. The US, electric shock, was administered through the grid floor. A commutator shifted the polarity of the grid bars four times per second. The shock was 550 v. ac applied through a variable current limiting resistor in series with *S*. The shock was continually regulated by *E* at 4.5 ma. Whenever *S* crossed the barrier, photocell beams were interrupted, a response was automatically recorded, and the trial terminated. Latencies of barrier jumping were measured from CS onset to the nearest .01 sec. by an electric clock. Stimulus presentations and temporal contingencies were controlled by automatic relay circuitry in a nearby room.

White masking noise at approximately 70-db. SPL was presented in both units.

Procedure.—The Escape group received escape training in the harness. Sixty-four unsignaled 6.0 ma. shocks were presented at a mean interval of 90 sec. (range, 60–120 sec.). If the dog pressed either panel with its head during shock, shock terminated. If the dog failed to press a panel during shock, shock terminated automatically after 30 sec. Two dogs were discarded for failing to escape 18 of the last 20 shocks.[4]

---

[4] It might be argued that eliminating these two dogs would bias the data. Thus naive

Twenty-four hours later dogs in the Escape group were given 10 trials of escape/avoidance training in the shuttle box: *S* was placed in the shuttle box and given 5 min. to adapt before any treatment was begun. Presentation of the CS began each trial. The CS-US interval was 10 sec. If *S* jumped the barrier during this interval, the CS terminated and no shock was presented. Failure to jump the barrier during the CS-US interval led to shock which remained on until *S* did jump the barrier. If no response occurred within 60 sec. after CS onset, the trial was automatically terminated and a 60-sec. latency recorded. The average intertrial interval was 90 sec. with a range of 60–120 sec. If *S* failed to cross the barrier on all of the first five trials, it was removed, placed on the other side of the shuttle box, and training then continued. At the end of the tenth trial, *S* was removed from the shuttle box and returned to its home cage.

The Normal control group received only 10 escape/avoidance trials in the shuttle box as described above.

The "Yoked" control group received the same exposure to shock in the harness as did the Escape group, except that panel pressing did not terminate shock. The duration of shock on any given trial was determined by the mean duration of the corresponding trial in the Escape group. Thus each *S* in the "Yoked" control group received a series of shocks of decreasing duration totaling to 226 sec.

Twenty-four hours later, *S*s in the "Yoked" control group received 10 escape/avoidance trials in the shuttle box as described for the Escape group. Seven days later, those *S*s in this group which showed the interference effect received 10 more trials in the shuttle box.

Results [5]

The Escape group learned to panel press to terminate shock in the harness. Each *S* in this group showed decreasing latencies of panel pressing over the course of the session ($p = .008$, sign test, Trials 1–8 vs. Trials 57–64). Individual records revealed that each *S* learned to escape shock by emitting a single, discrete panel press following shock onset. The *S*s in the "Yoked" control group typically ceased panel pressing altogether after about 30 trials.

TABLE 1

INDEXES OF SHUTTLE BOX ESCAPE/AVOIDANCE RESPONDING: EXP. I

Group	Mean Latency (in sec.)	% *S*s Failing to Escape Shock on 9 or More of the 10 Trials	Mean No. Failures to Escape Shock[a]
Escape	27.00	0	2.63
Normal Control	25.93	12.5	2.25
"Yoked" Control	48.22	75	7.25

[a] Out of 10 trials.

Table 1 presents the mean latency of shuttle box responding, the mean number of failures to escape shock, and the percentage of *S*s which failed to escape nine or more of the 10 trials during escape/avoidance training in the shuttle box for each group. The "Yoked" control group showed marked interference with escape responding in the shuttle box. It differed significantly from the Escape group and from the Normal control group on mean latency and mean number of failures to escape (in both cases, $p < .05$, Duncan's multiple-range test). The Escape group and the Normal control group did not differ on these indexes.

Six *S*s in the "Yoked" control group failed to escape shock on 9 or more of the 10 trials in the shuttle box. Seven days after the first shuttle-box treatment, these six *S*s received 10 further trials in the shuttle box. Five of them continued to fail to escape shock on every trial.

---

dogs which failed to learn the panel-press escape response in the harness might also be expected to be unable to learn shuttle box escape/avoidance. One of these dogs was run 48 hr. later in the shuttle box. It escaped and avoided normally. The other dog was too ill to be run in the shuttle box 48 hr. after it received shock in the harness.

[5] All p values are based upon two-tailed tests.

Discussion

The degree of control over shock allowed a dog during its initial exposure to shock was a determinant of whether or not interference occurred with subsequent escape/avoidance learning. Dogs which learned to escape shock by panel pressing in the harness did not differ from untreated dogs in subsequent escape/avoidance learning in the shuttle box. Dogs for which shock termination was independent of responding in the harness showed interference with subsequent escape learning.

Because the Escape group differed from the "Yoked" control group during their initial exposure to shock only in their control over shock termination, we suggest that differential learning about their control over shock occurred in these two groups. This learning may have acted in the following way: (*a*) Shock initially elicited active responding in the harness in both groups. (*b*) *S*s in the "Yoked" control group learned that shock termination was independent of their responding, i.e., that the conditional probability of shock termination in the presence of any given response did not differ from the conditional probability of shock termination in the absence of that response. (*c*) The incentive for the initiation of active responding in the presence of electric shock is the expectation that responding will increase the probability of shock termination. In the absence of such incentive, the probability that responding will be initiated decreases. (*d*) Shock in the shuttle box mediated the generalization of *b* to the new situation for the "Yoked" control group, thus decreasing the probability of escape response initiation in the shuttle box.

Escapable shock in the harness (Escape group) did not produce interference, because *S*s learned that their responding was correlated with shock termination. The incentive for the maintenance of responding was thus present, and escape response initiation occurred normally in the shuttle box.

Learning that shock termination is independent of responding seems related to the concept of learned "helplessness" or "hopelessness" advanced by Richter (1957), Mowrer (1960, p. 197), Cofer and Appley (1964, p. 452), and to the concept of external control of reinforcement discussed by Lefcourt (1966).

In untreated *S*s the occurrence of an escape or avoidance response is a reliable predictor of future escape and avoidance responding. Dogs in the "Yoked" control group and in the groups which showed the interference effect in Overmier and Seligman (1967) occasionally made an escape or avoidance response and then reverted to "passively" accepting shock. These dogs did not appear to benefit from the barrier-jumping—shock termination contingency. A possible interpretation of this finding is that the prior learning that shock termination was independent of responding inhibited the formation of the barrier-jumping—shock-termination association.

The *S*s in the "Yoked" control group which showed the interference effect 24 hr. after inescapable shock in the harness again failed to escape from shock after a further 7-day interval. In contrast, Overmier and Seligman (1967) found that no interference occurred when 48 hr. elapsed between inescapable shock in the harness and shuttle-box training. This time course could result from a temporary state of emotional depletion (Brush, Myer, & Palmer, 1963), which was produced by experience with inescapable shock, and which could be prolonged by being conditioned to the cues of the shuttle box. Such a state might be related to the parasympathetic death which Richter's (1957) "hopeless" rats died. Further research is needed to clarify the relationship between the learning factor, which appears to cause the initial occurrence of the interference effect, and an emotional factor, which may be responsible for the time course of the effect.

The results of Exp. I provide a further disconfirmation of the adaptation explanation of the interference effect. If *S*s in the "Yoked" control group had adapted to shock and, therefore, were not sufficiently motivated to respond in the shuttle box, *S*s in the Escape group should also have adapted to shock. Further, the Escape and the "Yoked" control groups were equated for the possibility of adventitious punishment for active responding by shock onset in the harness. Thus it seems unlikely that the "Yoked" control group failed to escape in the shuttle box because it had been adventitiously punished for active responding in the harness.

EXPERIMENT II

Experiment I provided support for the hypothesis that *S* learned that shock termination was independent of its responding in the harness and that this learning inhibited subsequent escape responding in the shuttle box. Experiment II investigates whether prior experience with *escapable* shock in the shuttle box will mitigate the effects of inescapable shock in the harness on subsequent escape/avoidance behavior. Such prior experience might be expected (*a*) to inhibit *S*'s learning in the harness that its responding is not correlated with shock termination and (*b*) to allow *S* to discriminate between the escapability of shock in the shuttle box and the inescapability of shock in the harness.

Method

Subjects.—The *S*s were 30 experimentally naive, mongrel dogs, weight, height, and housing as above. Three dogs were discarded: two because of procedural errors and one because of illness. The remaining 27 dogs were randomly assigned to three groups of nine *S*s each.

Apparatus.—The two units described for Exp. I were used.

Procedure.—The Preescape group received 3 days of treatment. On Day 1, each *S* received 10 escape/avoidance trials in the shuttle box as described in Exp. I. On Day 2, approximately 24 hr. after the shuttle-box treatment, each *S* in this group received an inescapable shock session in the harness. All inescapable shocks were unsignaled. The inescapable shock session consisted of 64, 5-sec. shocks, each of 6.0 ma. The average intershock interval was 90 sec. with a range of 60–120 sec. On Day 3, approximately 24 hr. after the inescapable shock, *S* was returned to the shuttle box and given 30 more escape/avoidance trials, as described for Day 1.

The No Pregroup received no experience in the shuttle box prior to receiving inescapable shock. On the first treatment day for this group, each *S* was placed in the harness and exposed to an inescapable shock session as described for the Preescape group, Day 2. Approximately 24 hr. later, *S* was placed in the shuttle box and given 40 trials of escape/avoidance training as described above. If *S* failed to respond on all of the first five trials, *S* was moved to the other side of the shuttle box. If *S* continued to fail to respond on all trials, it was put back on the original side after the twenty-fifth trial. Thus, if *S* failed to escape on every trial, it received a total of 2,000 sec. of shock.

The No Inescapable group was treated exactly as the Preescape group except that it received no shock in the harness. On Day 1, *S* received 10 escape/avoidance trials in the shuttle box. On Day 2, it was strapped in the harness for 90 min., but received no shock. On Day 3, it was returned to the shuttle box and given 30 more escape/avoidance trials.

Results

The No Pregroup showed significant interference with escape/avoidance responding in the shuttle box on Day 3. The Preescape and the No Inescapable groups did not show such interference. Figure 1 presents the mean median latency of jumping responses for the three groups (and a posterior control group, see below) over the four blocks of 10 trials. Analysis of variance on the three groups revealed that the effect of groups, $F\ (2, 24) = 3.55$, $p < .05$, and the effect of trial blocks, $F\ (3, 72) = 6.84$, $p < .01$, were significant. Duncan's multiple-range test indicated that the No Pregroup differed from the other two groups across all 40 trials both $p < .05$. The Preescape and the No Inescapable groups did not differ from each other. Similar results held

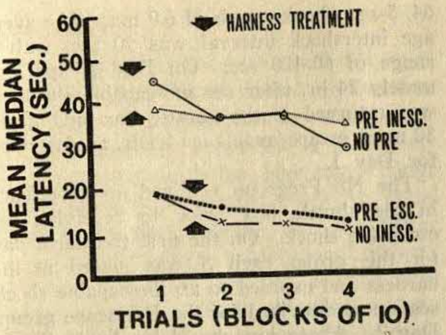


FIG. 1. Mean median latency of escape/avoidance responding. (The position of the arrow denotes whether the harness treatment occurred 24 hr. before the first or second block of trials.)

for the mean of mean latencies. A small, transitory disruption of improvement in shuttle-box performance following inescapable shock in the harness occurred in the Preescape group relative to the No Inescapable group. Difference scores for latencies between consecutive blocks of trials measure improvement in performance. A comparison of the Preescape group with the No Inescapable group on the difference between the mean latency on Trials 1–10 and the mean latency on Trials 11–20 revealed that the No Inescapable group showed significantly more improvement than the Preescape group, Mann-Whitney *U* test, $U = 15$, $p < .05$. No significant differences were found on difference scores for any subsequent blocks of trials.

Figure 2 presents the mean number of failures to escape shock for the three groups across the four blocks of trials. Analysis of variance revealed a significant overall effect of blocks, $F\ (3, 72) = 5.94$, $p < .01$, and a significant Groups × Blocks interaction, $F\ (6, 72) = 17.82$, $p < .01$. Duncan's tests indicated that the No Pregroup showed significantly more failures to escape than the other two groups across the 40 trials, both $p < .05$. The Preescape and the No Inescapable groups did not differ.

Figure 3 presents the total number of avoidance responses for the groups across the blocks of trials. Only the blocks effect was significant in the overall analysis of variance, $F\ (3, 72) = 27.90$, $p < .01$. No other effects were significant.

Panel presses made in the harness during the inescapable shock session

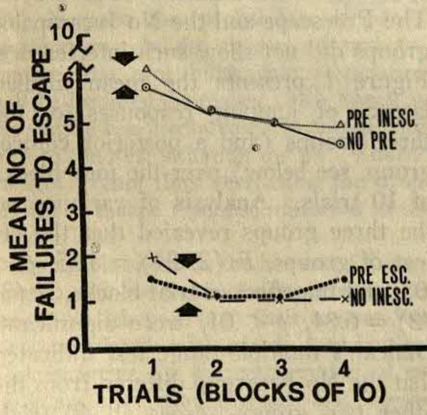


FIG. 2. Mean number of failures to escape shock. (The position of the arrow denotes whether the harness treatment occurred 24 hr. before the first or second block of trials.)

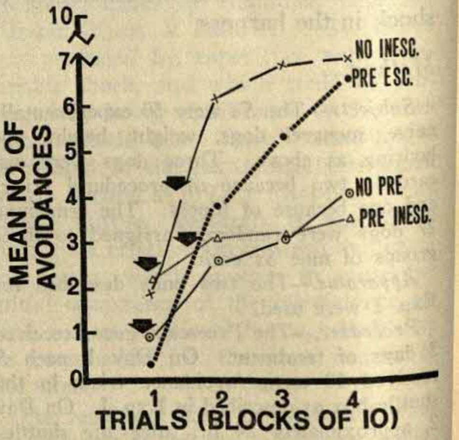


FIG. 3. Mean number of avoidances. (The position of the arrow denotes whether the harness treatment occurred 24 hr. before the first or second block of trials.)

were counted. On either side of *S*'s head were panels which *S* could press; panel pressing had no effect on the shock, but merely indicated attempts to respond and/or struggling in the harness. The Preescape group, having received 10 trials with *escapable* shock in the shuttle box the previous day, made more panel presses during the inescapable shock session than did the No Pregroup, the group for which the inescapable shock in the harness was the first experimental treatment, Mann-Whitney *U* test, $U = 9, p < .02$.

Posterior control group.—Subsequent to this experiment, a control group was run to determine if the *escapability* of shock in the shuttle box on Day 1 for the Preescape group was responsible for its enhanced panel pressing in the harness and lack of interference with responding in the shuttle box. Or would the mere occurrence of inescapable shock for a free-moving animal in the shuttle box have produced these results? Nine naive dogs received the following treatment: On Day 1, *S*s were placed in shuttle box and given 10 trials as for the Preescape and the No Inescapable groups. Unlike these groups, however, *S*'s barrier jumping did not (except adventitiously) terminate the shock and CS, because trial durations were programmed independently of *S*'s behavior. The duration of each of the 10 trials for this Preinescapable group corresponded to the mean trial duration for the Preescape and the No Inescapable groups on that trial. On Day 2, *S*s received 64 trials of inescapable shock in the harness. On Day 3, *S*s received 40 escape/avoidance trials in the shuttle box.

Figures 1, 2, and 3 present the escape/avoidance performance of the Preinescapable group on Day 3. In general this group performed like the No Pregroup. This impression was borne out by statistical tests. The Preinescapable group showed significantly slower median latency of barrier jumping than the Preescape and the No Inescapable groups across all 40 trials, both $p < .05$, Duncan's test. The Preinescapable group did not differ from the No Pregroup. Similar results held for the other indexes.

Analysis of the panel press data showed that the Preinescapable group made significantly fewer panel presses in the harness than the Preescape group, Mann-Whitney *U* test, $U = 14$, $p < .05$. The Preinescapable group did not differ significantly from the No Pregroup, $U = 26$.

Discussion

Three main findings emerged from Exp. II: (*a*) *S*s (Preescape), which first received escapable shock in the shuttle box, then inescapable shock in the harness, did not react passively to subsequent shock in the shuttle box, as did *S*s which either first received inescapable shock in the shuttle box (Preinescapable) or no treatment prior to shock in the harness (No Pre). (*b*) The Preescape group, having received experience with escapable shock in the shuttle box, showed enhanced panel pressing when exposed to inescapable shock in the harness, relative to naive *S*s given inescapable shock in the harness. Such enhanced panel pressing was specifically the result of the *escapability* of shock in the shuttle box: The Preinescapable group did not show enhanced panel pressing. (*c*) The interference effect persisted for 40 trials.

The *S*s which have had prior experience with *escapable* shock in the shuttle box showed more energetic behavior in response to inescapable shock in the harness. This contrasts with the interference effect produced by *inescapable* shock in *S*s which have had no prior experience with shock or in *S*s which have had prior experience with inescapable shock. Thus, if an animal first learns that its responding produces shock termination and then faces a situation in which reinforcement is independent of its responding, it is more persistent in its attempts to escape shock than is a naive animal.

General Discussion

We have proposed that *S* learned as a consequence of inescapable shock that its responding was independent of shock termination, and therefore the probability of response initiation during shock decreased. Alternative explanations might be offered: (*a*) Inactivity, somehow, reduces the aversiveness of shock. Thus *S* failed to escape shock in the shuttle box because it had been reinforced for inactivity in the harness. Since the interference effect occurred in *S*s which had been curarized during inescapable shock, such an aversiveness-reducing mechanism would have to be located inward of the neuro-myal junction. (*b*) *S* failed to escape in the shuttle box because certain responses which facilitate barrier jumping were *extinguished* in the harness during inescapable shock. In conventional extinction procedures, some response is first explicitly reinforced by correlation with shock termination, and then that response is extinguished by removing shock altogether from the situation. Responding during extinction is conventionally not *un*correlated with shock termination; rather, responding is correlated with the total absence of shock. In our harness situation, no response was first explicitly reinforced, and shock was presented throughout the session. A broader concept of extinction, however, might be tenable. On this view, any procedure which decreases the probability of a response by eliminating the incentive to respond is an extinction procedure. If the independence of shock termination and responding eliminates the incentive to respond (as assumed), then our harness procedure could be thought of as an extinction procedure. Such an explanation seems only semantically different from the one we have advanced, since both entail that the probability of responding during shock has decreased because *S* learned that shock termination was independent of its responses.

Learning that one's own responding and reinforcement are independent might be expected to play a role in appetitive situations. If *S* received extensive pretraining with rewarding brain stimulation delivered independently of its operant responding, would the subsequent acquisition of a bar press to obtain this reward be retarded? Further, might learned "helplessness" transfer from aversive to appetitive situations or vice versa?

If dogs learn in one situation that their active responding is to no avail, and then transfer this training to another shock situation, the opposite type of transfer (avoidance learning sets) might be possible: If a dog first learned a barrier-hurdling response which avoided shock in the shuttle box, would that dog be facilitated in learning to panel press to avoid shock in the harness (to a different CS)? Our finding, that dogs which first successfully escape shock in the shuttle box later showed enhanced panel pressing in the harness, is consonant with this prediction.

Does learning about response—reinforcement contingencies have its analogs in classical conditioning? If *S* experienced two stimuli randomly interspersed with each other (adventitious pairings possible), would it be retarded in forming an association between the two stimuli once true pairing was begun? Conversely, pretraining in which one stimulus is correlated with a US might facilitate the acquisition of the CR to a new CS. Pavlov (1927, p. 75) remarked that the first establishment of a conditioned inhibitor took longer than any succeeding one.

In conclusion, learning theory has stressed that two operations, explicit contiguity between events (acquisition) and explicit noncontiguity (extinction), produce learning. A third operation that is proposed, independence between events, also produces learning, and such learning may have effects upon behavior that differ from the effects of explicit pairing and explicit nonpairing. Such learning may produce an *S* who does not attempt to escape electric shock; an *S* who, even if he does respond, may not benefit from instrumental contingencies.

REFERENCES

BRUSH, F. R., MYER, J. S., & PALMER, M. E. Effects of kind of prior training and intersession interval upon subsequent avoidance learning. *J. comp. physiol. Psychol.*, 1963, **56**, 539–545.

COFER, C. N., & APPLEY, M. H. *Motivation: Theory and research.* New York: Wiley, 1964.

LEFCOURT, H. M. Internal vs. external control of reinforcement: A review. *Psychol. Bull.*, 1966, **65**, 206–221.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

OVERMIER, J. B., & SELIGMAN, M. E. P. Effects of inescapable shock on subsequent escape and avoidance learning. *J. comp. physiol. Psychol.*, 1967, **63**, 28–33.

PAVLOV, I. P. *Conditioned reflexes.* New York: Dover, 1927.

RICHTER, C. On the phenomenon of sudden death in animals and man. *Psychosom. Med.*, 1957, **19**, 191–198.

SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychol. Monogr.*, 1953, **67**(4, Whole No. 354).

(Received June 1, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 10–13

TEMPORAL PRESENTATION OF WINNING SYMBOLS AND SLOT-MACHINE PLAYING[1]

LLOYD H. STRICKLAND AND FREDERIC W. GROTE[2]
Carleton University, Ottawa
Dartmouth College

The effect of sequential presentation of winning symbols on the common slot machine was evaluated. It was found that Ss who saw winning symbols early in the outcome sequence were more likely to continue playing after a required number of trials than were those who observed a comparable proportion of winning symbols later in the sequence. The outcome was discussed in terms of S's expectations and in the context of "hope" and "disappointment" as these concepts are employed in Mowrer's (1960) learning theory.

This study was initiated as an extension of findings concerning sequential presentation of visual stimuli and their attributed frequency (Strickland, Katz, & Stifler, 1964). It was observed that the first occurring of two equally frequent light flashes was estimated to be the one which occurred more often, and this result was tentatively explained in the context of reinforcement delay (Kimble, 1962; Renner, 1964). The hypothesis tested here may have been long accepted by the manufacturers of the common "one-armed bandit," through the psychological literature does not reflect its investigation.

In the slot machine, three wheels spin when a coin is deposited and a lever pulled. They stop rotating in a 1–2–3 order, from left to right, at which point the player knows the outcome. The symbols exhibited in the machine's windows do not appear with equal frequency on the wheels: Wheel 1 usually has a relatively large proportion of potential winners (bells, oranges, etc.); these appear less often on Wheel 2 and rarely on Wheel 3. One may assume that the apparatus has been designed to maximize the likelihood of lever depression. Also, justification (beyond the vagaries of apparatus construction) may reasonably be inferred for two facts: (*a*) the array of outcomes is invariably presented sequentially, and (*b*) the "teaser" (e.g., a 2:1 return on a spray of cherries) is such only when this symbol stops on Wheel 1, while outcomes on Wheels 2 and 3 are irrelevant. The fact that the player is "instructed" about the location and reward value of these symbols by an itemized display attached to the machine seems to prohibit definition of these stimuli as conditioned reinforcers in the usual sense. Since the players of these devices are usually human, it is more meaningful to evaluate the hypothesis under consideration in terms of the effects of delayed knowledge of results (KR) or Mowrer's (1960) dis-

---

[1] This study was executed with partial support from the Filene Endowment in Human Relations, administered through Dartmouth College while the senior author was on the staff. Support for the junior author came from National Science Foundation Undergraduate Research Participation Grant G-12624. Funds for purchase and modification of the apparatus, costs for *S*s, etc., came from the Board of Christian Social Concerns of the Methodist Church, Washington, D. C. Appreciation for suggestions about the experiment is extended to T. K. Landauer, T. L. Tighe, and Charles L. Gruder.

[2] Now at the University of North Carolina, Chapel Hill.

cussion of "hope" and "disappointment" in learning contexts. Specifically, the hypothesis states that *S*s for whom winning symbols appear most frequently on the wheel that stops first should be more likely to continue playing after a series of "training trials" than those for whom the winning symbols appear most frequently on Wheel 3.

METHOD

A genuine one-armed bandit was used. Upon activation, Wheel 1 ceased rotation after 4.5 sec. with .2-sec. intervals separating this from the stopping of Wheel 2 and 2 from 3. One thousand plays demonstrated that the machine was *mechanically fair*—no wheel stopped more frequently at one position than at another, and no interwheel contingencies could be discerned. The usual symbols were replaced with randomly placed colored bars. Wheel 1 had 14 red and 6 green bars; Wheel 2 had 10 of each color; Wheel 3 had 6 red and 14 green. Thus, specification to *S* of the winning color would determine whether he observed frequent winning symbols "early" or "late" in the sequence, without altering the apparatus or the overall probability (.105) of obtaining three identical bars. To control for potential bar-color effects, the displays on Wheels 1 and 3 were interchanged for half the *S*s. A pen recorder was connected to the machine, so that *S*'s rate of play could be measured.

The *S*s were 44 male students with a median age of 15.9 yr. A tape recorder presented the instructions, identical for all *S*s except for specification of the winning color. It was made clear that (*a*) *S* must play all of the 100 nickels with which he was presented by *E*; (*b*) every time *S* got three red (green) bars, 40¢ would be dropped in a cup adjoining the payoff tray by *E*; (*c*) after the 100 trials *S* could continue playing with his winnings or not as he chose, stopping at any time and keeping 50% of his holdings. On each trial, *S* was required to read to *E* the outcomes on the wheels to control for attention factors—it was felt that if *S* saw a sure "loser" appear on Wheel 1, he might not attend to other outcomes and lend support to the hypothesis for reasons ascribable to *apparent* partial reinforcement (e.g., Lewis & Duncan, 1957, 1958). The machine's payoff mechanism was not employed, so that its "autonomy" was minimized; *S* always received 40¢ for winning, and the fact that such machines retain and conceal a percentage for the "house" whenever payoff tubes are full was of no consequence. This amount, paid out on 10.5% of the trials, was such as to discourage continued play by the "rational operator," i.e., the longer he played, the less money he had. Two groups were presented the 70–50–30 sequence, the other two the 30–50–70 sequence. Half the *S*s in each group won on red, the other half on green.

TABLE 1

CHARACTERISTICS OF APPARATUS AND *S*S

Group	Percentage Winning Symbols seen by *S*s on Each Wheel—Trials 1–100			Mean No. Trials on which *S*s Won	Mean Trial on which *S*s Won[a]
	Wheel 1	Wheel 2	Wheel 3		
I	70.5	51.1	30.9	12.9	50.4
II	69.6	50.3	29.6	10.1	51.5
III	32.0	48.5	71.6	10.6	49.9
IV	30.5	49.6	68.1	10.6	52.8

[a] Mean of *S*s means.

RESULTS

In Table 1 it may be seen that the slot machine behaved appropriately for all groups. It is apparent that the proportions of winning symbols observed approximated closely those intended; the four groups won at a rate close to the theoretical probability; the four groups' mean winning trials did not differ from the mean (50.5) of the theoretical, rectangular distribution of winning trials.

Data bearing on the hypothesis appear in Table 2. Since no difference attributable to color of symbols appeared, the data from the two replications were combined in a fourfold table with symbol sequences (70–50–30 or 30–50–70) as one classification and "continue" or "stop" as the other. A χ^2 value, corrected for continuity, of 4.61 ($p < .05$, two-tailed test) was computed.

TABLE 2

Dependent Measures

Measures	Groups I & II	Groups III & IV
Proportion of *S*s Continuing to Play after Training Trials	$\frac{7+6}{22}$	$\frac{2+3}{22}$
Rate of Play during Training Trials		
Minutes	.151	.148
No. *S*s	18	21
Mean Trials Exceeding 100 for *S*s Electing to Continue	16.1	16.8

It was ascertained that differential machine behavior was not reflected by those stopping or continuing. Both those who continued playing and those who stopped received their wins near the theoretical mean trial (50.5). For the former, the mean was 50.3, for the latter, 50.0. The former group did not achieve more wins over the first 100 trials (10.9) than did the latter (11.2). Within each of the four treatments, *S*s who made more than the theoretical number of wins were not more likely to continue than those who made less. It was ascertained that 13 of the 22 *S*s in Groups I and II saw the machine as "fair" after the first 100 trials, as did 11 *S*s of Groups III and IV; there was no indication that this perception was related to continued playing.

Faulty pen-recorder function led to slightly reduced *n*'s in Groups I, II, and III, but there were no differences in rate of play (Table 2) among the remaining *S*s. This was not unexpected—the apparatus, with its built-in *minimum* 5-sec. intertrial interval, is an imperfect one for measuring this variable.

The groups did not differ in terms of the number of trials which they continued to play with their own money (Table 2). This was anticipated, since at the point where the decision was made to continue, a variety of cognitive rules could be (and, in fact, were) formed consistent with the decision. Postexperimental questionnaires yielded verbal justifications (e.g., "I'll play as long as I can keep $2.00"; "I'll play 8 nickels and if I don't win I'll stop," etc.) which did not appear to be different for those continuing play in the two major conditions. Genuine extinction trials could not be run without blatent modification of the machine.

Attempts to assess the role of delayed KR have frequently been confounded by manipulation of intertrial interval (e.g., Bilodeau & Bilodeau, 1958; Denny et al., 1960). The present apparatus appears to avoid such enforced confounding, with its constant temporal and sequential presentation of final outcomes, but the *pattern* of KR varied drastically for the several groups. In Groups I and II, *S*s could entertain the covert hypothesis, "I might win" on 70% of the trials after the first wheel has stopped; in III and IV this proportion is only 30%. Comparable proportions are 35% and 15% after two wheels have stopped. It seems that early, frequent, and extended anticipation of a win promotes continued playing. Assuming that *S*s want to win money, these percentages may be interpreted in terms of frustration, a variable of clear relevance (Amsel, 1958) for reinforcement-delay studies. Groups III and IV may be initially frustrated more often, but when they see winners in the first two wheels, a winner on the third approaches (.7) a "sure thing" and disconfirmation may be quite acute affectively. Groups I and II are more often "strung along" by the outcomes on Wheels 1 and 2, but a winner on Wheel 3 is relatively (.3) unlikely, may not be expected, and disconfirmation may have less impact.

Discussion

Discussion of hypotheses about "might win" and frustration leads to consideration of the concepts of hope and disappointment as discussed by Mowrer (1960). In terms of his theory, the winning stimuli used here may be seen as Type-2 secondary reinforcers which refer "... to the reward or satisfaction experienced when a stimulus comes *on*, indicating the imminent occurrence of some *desired* event... secondary reinforcement type-2 implies *hope* (presentation of a promise) [Mowrer, 1960, p. 101; italics his]." He also states, "Clearly enough, if a signal has acquired the capacity to arouse hope... and if such a stimulus is presented and withdrawn before it is confirmed there will be a recrudescence of... hopelessness. The individual, as one may say, 'loses hope,' 'feels disappointed' [p. 164]." Presentation and withdrawal of winning symbols on the spinning wheels certainly occurs. On this basis, one may expect that more *S*s should remain hopeful in Groups I and II than in Groups III and IV, if proximity of Type-2 secondary reinforcement to the response is important.

Although these intervening variables may have intuitive appeal, their use in an objective analysis of slot-machine playing may be questionable. Perhaps their employment is most justified in consideration of initial acquisition of the slot-machine "habit," where verbal mediation by *S* might be most frequent. It will be remembered that the present *S*s were unsophisticated high-school age students. They did not keep the machine running at its maximum rate, and it is presumed that their behavior differed from that of the "addicts" who haunt the casinos and, in an apparently mechanical manner, keep a machine running constantly. It is possible that the maintenance of the latter's high-response rate may be looked at in terms of partial reinforcement and differential delay of conditioned reinforcers without recourse to intervening variables. Unsystematic observation leads to the conclusion that overt signs of emotionality and superstitious behavior are less frequent (have been extinguished) among the veterans than among our neophyte *S*s.

At any rate, the present finding suggests slot-machine playing is not only a function of attention or of the partial-reinforcement schedule of the payoffs (e.g., Lewis & Duncan, 1957, 1958). The temporal distribution of the "win" symbols across the three wheels is apparently not a chance outcome of the machine's design.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Bilodeau, A. E., & Bilodeau, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

Denny, M. R., Allard, M., Hall, E., & Rokeach, M. Supplementary report: Delay of knowledge of results, knowledge of the task, and intertrial interval. *J. exp. Psychol.*, 1960, **60**, 327.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1961.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, **54**, 115–120.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

Renner, E. K. Delay of reinforcement: A historical review. *Psychol. Bull.*, 1964, **61**, 341–361.

Strickland, L. H., Katz, A. E., & Stifler, L. T. P. Sequential order and perceived frequency of occurrence of equally frequent stimuli. *Percept. mot. Skills*, 1964, **19**, 795–801.

(Received April 15, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 14–19

ACQUISITION OF PAIRED ASSOCIATES AS A FUNCTION OF ASSOCIATIVE STRENGTH OF COMPETING RESPONSES[1]

RICHARD BLOCH AND JAMES F. VOSS

University of Pittsburgh

8 modified paired associates were presented in Experiment I for 20 trials. One response, R_1, occurred with a particular stimulus on every other trial. The R_2 response, alternating with the R_1 response, was systematically varied. In Cond. SA, MA, WA, and NA, R_2 was a strong associate, moderate associate, weak associate, or nonassociate of the respective stimulus. A non-S-R_2 condition also was employed. In Experiment II, the R_1 responses were nonassociates or associates of the respective stimuli and the R_2 responses were associates, nonassociates, or S-R_2 pairs not presented. The correct response and intrusion results indicated that acquisition of nonassociated S-R_1 pairs was poorest when the S-R_2 pair also was nonassociated. Moderate S-R_2 associates did not interfere significantly with S-R_1 nonassociate acquisition. When S-R_1 was an associated pair, interference occurred whether the respective S-R_2 pair also was associated or nonassociated. The results suggested that competition of responses of similar associative strength produce interference, especially for weak associations, and also suggested that strong associates are inhibited in the acquisition of new associations.

The purpose of the two experiments reported in this paper was to determine how acquisition of one association (S-R_1) is influenced by the associative strength of a concomitantly presented second association (S-R_2). Experiment I consisted of a 20-trial paired-associate recall task in which S-R_1 and S-R_2 were presented on alternating trials. The S-R_1 pairs were nonassociated words and were identical for each of five S-R_2 experimental conditions; the R_2 responses, however, were strong, moderate, weak, or nonassociates of S. In a fifth condition, no S-R_2 pairs were presented. In Exp. II, associates or nonassociates were employed as S-R_1 pairs, and the responses of the S-R_2 pairs were associates, nonassociates, or blanks (no responses). Thus, the design of Exp. I provided for determination of the influence of competing response associative strength upon the acquisition of new associations; the design of Exp. II provided for determination of the influence of associates and nonassociates upon the learning of associates and nonassociates.

EXPERIMENT I

Method

Procedure.—Eight paired associates were presented via recall method for 20 trials at a 3-sec. pairing and 3-sec. testing rate with a 6-sec. interval between eight pairings and eight stimulus presentations. Two responses presented with each stimulus alternated on a trial-by-trial basis: S-R_1 occurred on the odd-numbered trials and S-R_2 on the even-numbered trials for half of the *S*s; for the remaining half of the *S*s, S-R_1 and S-R_2 were reversed with respect to odd or even trials. Thus, each S-R_1 and each S-R_2 association occurred 10 times on an alternating trial basis.

The eight S-R_1 associations were identical for each of five experimental conditions. The S-R_2 associations were varied for five groups: Cond. SA—the R_2 words were strong associates of the respective stimuli; Cond. MA—the R_2 words were moderate

---

[1] The research reported in this paper was supported by the National Institute of Child Health and Human Development (HD-00957-03).

TABLE 1

VERBAL MATERIALS OF EXP. I AND II

EXP. I

S	R_1	R_2				
		SA	MA	WA	NA	BL
Bread	Doctor	Butter	Food	Hunger	Dark	—
Cold	Lock	Hat	Warm	Hard	Dream	—
Fruit	Iron	Apple	Fly	Basket	Walk	—
Lamp	Road	Light	Shade	Floor	Pen	—
Ocean	Glass	Water	Sea	Green	Bird	—
Shoes	Knife	Feet	Socks	Lace	Head	—
White	Bush	Black	Snow	Bear	Smile	—
Scissors	Bank	Cut	Sharp	Steel	Table	—

EXP. II

S	R_1		R_2		
	A	NA	A	NA	BL
Cheese	Mouse	Visit	Bread	Ride	—
Earth	Ground	March	Dirt	Heart	—
Foot	Hand	Table	Leg	Bank	—
Moon	Sun	Smile	Night	Bush	—
Window	Pane	Knife	Door	Dream	—
Baby	Boy	Lead	Cry	Post	—
Needle	Sew	Lock	Sharp	Fire	—
Ocean	Blue	Bird	Sea	Glass	—

associates; Cond. WA—the R_2 words were weak associates; Cond. NA—the R_2 words were nonassociates[2]; Cond. BL—no R_2 presentations occurred.

Apparatus and materials.—The paired associates were presented via a Carousel projector. Table 1 presents the materials employed. The stimuli and responses were selected from the Palermo and Jenkins (1964) norms. With four exceptions, all responses were rated as A or AA by Thorndike-Lorge (1944) norms. Four random orders for the pairings and four different random orders for the test stimuli were employed.

Subjects.—The *N*/each of the five conditions was 12, with half of the *S*s of each condition assigned to each of the two alternation conditions. The *S*s were students satisfying an introductory psychology requirement and were assigned to all conditions via a table of random numbers. Recall instructions were presented which included instructions to recall the last response which was presented with the respective stimulus.

Results and Discussion

S-R_1 correct responses.—The second column of Table 2 presents mean correct S-R_1 responses and *SD*s as a function of S-R_2 condition. The data indicate that mean correct S-R_1 responses were greatest for Cond. BL, next largest for the MA and SA conditions, and least for the WA and NA conditions. Since Cond. BL served as a base-line control, the Dunnett test (Edwards, 1960) was employed to determine which conditions, if any, yielded performance significantly poorer than the control (BL) condition. Condition BL performance was significantly

---

[2] The question of whether a nonassociate initially has a zero associative strength is open to debate. The present writers would contend that any two words in the repertoire of an individual have an associative strength > 0.

TABLE 2

MEAN CORRECT R_1 AND R_2 RESPONSES FOR EXP. I

$S\text{-}R_2$ Cond.	Mean Correct R_1 Rs		Mean Correct R_2 Rs	
	M	σ	*M*	σ
SA	60.6	9.57	78.7	1.18
MA	66.2	10.35	76.7	3.55
WA	58.8	9.59	73.0	4.36
NA	55.8	14.35	64.4	10.86
BL	67.4	7.56	—	—

superior to that of Cond. SA, t (50) = 4.76, $p < .01$, Cond. WA, t (50) = 6.01, $p < .01$, and Cond. NA, t (50) = 8.11, $p < .01$. The t value of the comparison of Cond. BL and MA was < 1.00. These findings indicate that the $S\text{-}R_2$ associates of Cond. SA, WA, and NA interfered to a significant degree with the acquisition of the $S\text{-}R_1$ associations. An analysis of variance performed upon the correct R_1 response data also revealed that the $S\text{-}R_1$, $S\text{-}R_2$ alternation condition source of variation is not significant, F (1, 50) = 1.24, nor does alternation interact significantly with $S\text{-}R_2$ condition, F (4, 50) < 1.00.

Intrusions.—The top half of Table 3 presents a summary of the $S\text{-}R_1$ intrusion frequency data. Columns 2–7, respectively, indicate total intrusions, intrusions from the R_2 word associated with the same stimulus word, intrusions from other R_2 words of the list, intrusions from other R_1 words, intrusions from stimulus words, and extralist intrusions. The most important finding of the data is that intrusions were greatest in the NA $S\text{-}R_2$ condition and that the major component of these intrusions was the intrusions from the associated $S\text{-}R_2$ word. The intrusion data thus support the view that acquisition of $S\text{-}R_1$ in Cond. NA was inferior because of interference produced by the NA $S\text{-}R_2$ associations. Similarly, in Cond. WA, intrusion frequency was relatively great although the percentage of intrusions from R_2 were less in Cond. WA (38.9%) than in Cond. NA (69.9%). It also is of interest to note that the percentage of R_2 intrusions in Cond. SA (30.0%) exceeded the corresponding percentage of Cond. MA (19.5%).

TABLE 3

SUMMARY OF INTRUSION DATA OF EXP. I

$S\text{-}R_2$ Cond.	Total Int.	R_2	Other R_2	Other R_1	Stim.	Extra-list
$S\text{-}R_1$						
SA	40	12	4	13	10	1
MA	41	8	8	14	7	4
WA	72	28	6	27	9	2
NA	146	102	9	27	6	2
BL	32	0	0	11	17	4
$S\text{-}R_2$						
SA	8	5	1	0	2	0
MA	18	11	3	2	0	2
WA	39	14	7	9	7	2
NA	84	61	11	3	6	3

$S\text{-}R_2$ correct response data.—The mean correct response frequencies of the $S\text{-}R_2$ associations are presented in Column 3 of Table 2. An analysis of variance revealed a significant $S\text{-}R_2$ condition source of variation, F (3, 40) = 13.63, $p < .01$. Comparison of each pair of conditions by use of Scheffé's test (Edwards, 1960) indicated that each condition differed significantly from each other condition at the .01 level with the exception that Cond. SA and MA do not yield a significant difference in correct R_2 responses, F (3, 40) = 6.86 with a .05 F' requirement of 8.52. Thus, these data indicate that, in general, performance increases as strength of association increases.

Analysis also revealed a significant order effect F (1, 40) = 5.14, $p < .05$, in $S\text{-}R_2$ acquisition, with superior performance occurring when $S\text{-}R_2$ associations were presented on the ood-numbered trials and S-R_1 associations on the even-numbered trials. Although the Cond. × Order interaction is not significant, $F\ (3, 40) = 2.20$, it should be mentioned that the greatest difference in order occurred in Cond. NA. This finding supports the previous results of Popp and Voss (1965).

S-R_2 intrusion Data.—The lower half of Table 3 presents the S-R_2 intrusion frequency data. The most important aspect of the data is the relatively large number of intrusions found in Cond. NA, with the data indicating that the largest component of this total was the R_1 word associated with the same stimulus (72.6%).

EXPERIMENT II

Experiment II was designed to extend the study of the influence of S-R_2 associative strength upon S-R_1 acquisition by employing associates as S-R_1 pairs as well as nonassociates and determining how acquisition of associated or nonassociated S-R_1 pairs varies as a function of associated, nonassociated, or blank S-R_2 conditions.

Method

Procedure.—The procedure in Exp. II was similar to that of Exp. I except for the experimental groups and word pairs. The Exp. II design consisted of two S-R_1 conditions, associates (A_1) or nonassociates (NA_1). Each S-R_1 condition was presented under three S-R_2 conditions, associates (A_2), nonassociates (NA_2), or no S-R_2 associations (BL) in which no stimulus or response occurred on a S-R_2 trial. Table 1 presents the verbal materials employed in Exp. II. The respective R_1 and R_2 associated words were chosen from word norms (Jenkins & Palermo, 1964) so that their strengths, according to the norms, were roughly equivalent.

Results

TABLE 4

MEAN CORRECT R_1 RESPONSES FOR EXP. II

S-R_2 Cond.	S-R_1 Cond.			
	A_1	σ	NA_1	σ
A_2	71.3	6.64	70.3	7.04
NA_2	74.9	4.56	59.8	12.96
BL	79.0	1.15	72.5	4.96

Correct S-R_1 response data.—Table 4 presents the mean correct response frequencies and *SD*s for the A_1 and NA_1 conditions under S-R_2 A_2, NA_2, and BL conditions. The data indicate that S-R_2 associates interfered more than S-R_2 nonassociates in the acquisition of S-R_1 associates and that S-R_2 nonassociates interfered more than S-R_2 associates in the acquisition of S-R_1 nonassociates. A Dunnett test of the A_1 data revealed that Cond. BL performance was significantly superior to Cond. A_2 performance, $t\ (30) = 6.61$, $p < .01$ and to Cond. NA_2 performance, $t\ (30) = 5.81$, $p < .01$. Analysis of variance also revealed no significant effect of S-R_1, S-R_2 order ($F < 1.00$) and no significant Cond. × Order interaction, $F\ (2, 30) = 2.04$. With respect to the NA_1 data, Cond. BL yielded performance significantly superior to Cond. NA_2, $t\ (30) = 3.24$, $p < .01$, but there is no significant difference in Cond. BL and Cond. A_2 performance, $t\ (30) < 1.00$. Analysis of variance also revealed no significant order effect ($F < 1.00$) and no significant Cond. × Order interaction ($F < 1.00$).

Intrusion data.—Table 5 presents the intrusion frequency data. The S-R_1 A_1 data indicate that a relatively large frequency of intrusions occurred in the S-R_2 A_2 condition and that most of these intrusions (82.7%) were due to the competing S-R_2 association. For the S-R_1 NA_1 condition, intrusion frequency was greatest for the NA_2 con-

TABLE 5

SUMMARY OF INTRUSION DATA OF EXP. II

$S\text{-}R_2$ Cond.	Total Int.	R_2	Other R_2	Other R_1	Stim.	Extralist
$S\text{-}R_1$ A_1						
A_2	75	62	—	7	2	4
NA_2	20	9	—	8	3	—
BL	4	—	—	2	1	1
$S\text{-}R_1$ NA_1						
A_2	29	6	4	15	2	2
NA_2	79	48	7	22	2	—
BL	33	—	—	22	9	2

dition. Again, the majority of these intrusions came from the competing response (60.8%).

DISCUSSION

The major result of the present experiment is that in the acquisition of nonassociated $S\text{-}R_1$ pairs, interference is maximal when the competing $S\text{-}R_2$ pairing is also nonassociated. Furthermore, the findings suggest that interference attributable to the competing $S\text{-}R_2$ associations decreases as strength of the $S\text{-}R_2$ association increases. With respect to the acquisition of associated $S\text{-}R_1$ pairs, the results of Exp. II suggest associated as well as nonassociated $S\text{-}R_2$ pairs interfere significantly with the associated pair interfering slightly more than the nonassociated.

Considering first the $S\text{-}R_1$ nonassociates, two hypotheses are suggested to account for these findings: (*a*) a competing response hypothesis, namely, acquisition is a function of the similarity of $S\text{-}R_1$ and $S\text{-}R_2$ associative strengths such that similar associative strengths produce more competition and poorer acquisition whereas an increase in the difference of $S\text{-}R_1$ and $S\text{-}R_2$ associative strengths improves acquisition; (*b*) an inhibition hypothesis, namely, acquisition of one association, $S\text{-}R_1$, decreases when $S\text{-}R_2$ associative strength is relatively large because *S* is required to inhibit the stronger response, R_2, in order to acquire $S\text{-}R_1$. The first hypothesis provides an explanation for the finding that acquisition of $S\text{-}R_1$ was poorest for the $S\text{-}R_2$ nonassociates, next poorest for the weak associates, and, finally, why moderate associates yielded no significant interference in either experiment. The first hypothesis does not explain, however, why strong $S\text{-}R_2$ associates produced interference in $S\text{-}R_1$ acquisition in Exp. I. The second hypothesis suggests that inhibition of the strong R_2 responses occurs in the acquisition of $S\text{-}R_1$, and this inhibition process, likely taking place on the $S\text{-}R_1$ trials, interferes with $S\text{-}R_1$ acquisition. It may be noted that the intrusion data are consistent with both hypotheses.

Considering the $S\text{-}R_1$ associates, it may be assumed that because the associative strength initially is substantial, $S\text{-}R_1$ is less susceptible to certain interference effects, e.g., those of other responses of the list. The intrusion data do indicate, however, that competition again does occur when $S\text{-}R_2$ is an associate and this interference produces a significant decrement in performance. Why the $S\text{-}R_2$ nonassociate also produced a significant decrement is not apparent, although the intrusion data show relatively few $S\text{-}R_2$ nonassociate intrusions.

The present results extend the findings of Popp and Voss (1965) and of probabilistic verbal acquisition in general. Popp and Voss presented $S\text{-}R_1$, $S\text{-}R_2$ associates for 40 trials with $P(S\text{-}R_1) = P(S\text{-}R_2) = .50$ and varied $S\text{-}R_1$, $S\text{-}R_2$ pattern. Although the study differed from the present experiments in that the anticipation method, low frequency words, and six S-R_1, S-R_2 associates were used, two points of comparison may be noted. First, in the probabilistic situation, the recall method enables *S* to make a correct response on each trial. On the other hand, the anticipation method does not effectively provide such an opportunity with a random R_1-R_2 pattern and only permits correct responses on each trial of a structured R_1-R_2 pattern if *S* understands the pattern. It is likely, therefore, that in the S-R_1, S-R_2 situation in which S-R_1 is nonassociated and S-R_2 is associated, a relatively greater number of R_2 responses would occur with the anticipation method than in the present study unless *S* understood that the pattern was one of alternation. Second, probabilistic verbal experimentation typically has involved an attempt to initially equate S-R_1 and S-R_2 associative strengths and study performance as a function of S-R_1, S-R_2 probability-ratio. The present experiments extend the study of S-R_1, S-R_2 associates to a condition in which the associative strengths of S-R_1 and S-R_2 are varied initially.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

PALERMO, D. S., & JENKINS, J. J. *Word association norms.* Minneapolis, Minn.: Univer. Minnesota Press, 1964.

POPP, R., & VOSS, J. F. Acquisition-transfer and retention of S-R_1, S-R_2 associations as a function of S-R_1, S-R_2 pattern. *J. exp. Psychol.,* 1965, **70,** 304–308.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* Teachers College, Columbia University, Bureau of Publications, 1944.

(Received April 27, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 20–25

DIFFERENTIAL GSR CONDITIONING AT EXTENDED INTERSTIMULUS INTERVALS[1]

DARWIN P. HUNT

University of Dayton

6 groups of 12 Ss each were given 24 differential GSR conditioning trials followed by 6 extinction trials to the previously reinforced CS. The S_+, a 1,000 or 1,967-cps. tone in a counterbalanced fashion, was followed by an electric shock; the S_- was the remaining tone and was not followed by the shock. Each S received 12 S_+ and 12 S_- trials in a random order during acquisition. The treatment of the 6 groups differed in terms of the interstimulus interval (ISI) employed during acquisition: ½, 1, 2, 3, 4, or 5 sec. It was found that the maximum differential response (DR) was associated with the 4-sec. ISI and was significantly greater than the DR at either the ½ or 5-sec. ISI. The greatest CR to the S_+ occurred at the 1-sec. ISI. Extinction of the response to the S_+ was not related to the ISI. A primitive model to account for some of the findings is described.

An interstimulus interval (ISI) of approximately ½ sec. is typically indicated as the optimal ISI for simple conditioning (e.g., Kimble, 1961). The possibility that an ISI longer than ½ sec. would result in improved differential conditioning has been suggested by Gynther (1957). In a study of differential eyelid conditioning, he found that "even under the most effective conditions for the development of discrimination, a considerable percentage of responding (to the nonreinforced stimulus) persists (p. 414)." He had used a ½ sec. ISI and offered the possible explanation that this time was too short to allow differentiating verbal reactions to occur.

This general conjecture was supported by the finding of Hartman and Grant (1962) that the magnitude of the differential response (DR) in eyelid conditioning was positively related to the duration of the ISI over the range of intervals investigated (400 to 1,000 msec.). It is of interest to note that a more detailed analysis in which *S*s were separated into two groups, conditioners and voluntary responders, revealed that the conditioners exhibited an optimum ISI at 800 msec.; the voluntary responders showed a monotonic improvement in DR with increased ISI. Kimmel and Pennypacker (1963) studied the effects of ISI, over a range of ¼–2 sec., on differential GSR conditioning and also report a monotonic and direct relationship between the DR and ISI.

Based on the above and a consideration of the general conditioning literature, a primitive descriptive model of the differential conditioning process was constructed. Regarding the development of the CR_+, i.e., the CR to the reinforced CS (S_+), one would expect some optimum range of ISI to be evidenced (e.g., Kimble, 1961). As a first approximation, the relationship between the CR and the ISI which seems to apply to a simple conditioning situation is assumed (Fig. 1a). With regard to the relationship between the CR_-, i.e., the CR to the nonreinforced CS (S_-), and the ISI one might expect that the generalized CR_+ to the S_-

---

[1] This investigation was supported in part by Research Grant MH 08433 from the Public Health Service, National Institutes of Health. The assistance of Howard Stevens and Philip Piro is appreciated.

would result in an increased CR_- over the shorter intervals; and extensions of the ISI beyond some minimal duration, say $\frac{1}{2}$ sec., would increasingly permit some discriminative reaction, e.g., verbal mediational response, to be reflected. Thus the CR_- function might be of the same general form as the CR_+ function, but the rate at which it decreases at the longer ISI would be greater and the interval at which an essentially zero response occurs would be shorter (Fig. 1a).

In this primitive differential conditioning model the DR is reflected by the discrepancy between the CR_+ and the CR_- functions. A plot of these discrepancies as a function of the ISI is shown in Fig. 1b. It is this hypothesis, i.e., for differential GSR conditioning there is an optimum ISI at some interval longer than $\frac{1}{2}$ sec., to which the study reported below addresses itself.

METHOD

Subjects.—At the encouragement of their instructors, 97 introductory psychology students at the University of Dayton served as *S*s. The data of 25 *S*s were eliminated due to procedural errors, excessive base-level skin resistances, apparatus failure, etc.

Apparatus.—The CSs were a 1,000 and a 1,967 cps. tone adjusted to a loudness of 1 sone. These tones were generated by fixed frequency oscillators (M.F. Electronics Corporation) and were presented to *S* through the right earphone of a MSA Noisefone, Mark II, headset. To avoid a click, the rise time of all tones was adjusted to 25 msec. using a Grason-Stadler electronic switch, Model 829 C. The UCS was a DC electric shock modulated by a Hewlett-Packard square wave generator, Model 2HA, at 50 cps. with equal on-off times. The shock electrodes were placed on the thumb and middle finger of the right hand. A constant current generator was designed and constructed so that current levels from 1.5 to 2.5 ma. could be selected in .1-ma. steps. The duration of the UCS was 100 msec. although the shock was delivered as five 10-msec. pulses separated by 10-msec. intervals.

The durations of the CS and the UCS as well as the ISI were controlled by Tektronix waveform and pulse generators. A Systron counter and timer, Model 1031, was employed to check the frequencies of tones and to calibrate the durations of the CS, the UCS, and the ISI. The *S*'s skin resistance was measured with a Fels dermohmmeter, Model 22A, using the electrodes furnished with the dermohmmeter and the recommended electrode paste. A special spring-loaded guillotine device was constructed to hold the electrodes firmly to the palm and the back of *S*'s left hand.

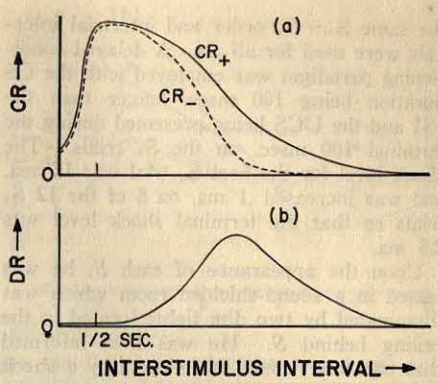


FIG. 1. Primitive model of differential conditioning depicting (*a*) the response to S_+ and S_- and (*b*) the differential response, as a function of the interstimulus interval.

Procedure.—Twelve *S*s were run under each of six ISI conditions: $\frac{1}{2}$, 1, 2, 3, 4, or 5 sec. The S_+ was either the 1,000- or the 1,967-cps. tone with the other tone serving as the S_-. Three *S*s using each S_+ were tested by each of two *E*s for each of the six ISI conditions. Each *S* was given 12 S_+ and 12 S_- trials during acquisition. A nonreinforced test trial to the S_+ was administered following the twelfth trial. The differential conditioning trials were followed immediately by six extinction trials to the S_+.

The S_+ and S_- trials were administered in a random order with the restrictions that (*a*) the first and twenty-third acquisition trial must be S_+, (*b*) the twelfth and twenty-fourth acquisition trial must be S_-, (*c*) no more than two S_+ or S_- trials could occur consecutively, and (*d*) the first 12 acquisition trials must contain an equal number (6) of S_+ and S_- trials. The intertrial intervals, measured from the successive onsets of the CSs, were 30, 40, 50, or 60 sec. ± 5 sec. These intervals were assigned nonsystematically so that each of the four ITIs occurred approximately equally often. For all *S*s the interval preceding the test trial and Trials 12, 24, and 25 was 40 sec. Once established, the same $S_+ - S_-$ order and intertrial intervals were used for all *S*s. A delayed conditioning paradigm was employed with the CS duration being 100 msec. longer than the ISI and the UCS being presented during the terminal 100 msec. on the S_+ trials. The shock level for the first S_+ trial was 1.7 ma. and was increased .1 ma. on 8 of the 12 S_+ trials so that the terminal shock level was 2.5 ma.

Upon the appearance of each *S*, he was seated in a sound-shielded room which was illuminated by two dim lights located in the ceiling behind *S*. He was then informed that one tone would be followed by a shock and the other tone would not, and that he should relax but keep his eyes open. After attaching the headset, the GSR electrodes and the shock electrodes, the S_+ tone assigned to that particular *S* was presented alone three times at 30-sec. intervals. Then three shocks were delivered at 30-sec. intervals at successive shock levels of 1.5, 1.6, and 1.7 ma. A base-level skin resistance was determined 70 sec. after the third pretest shock. The appropriate acquisition and extinction trials were then administered. Two skin resistance measures were read directly from the meter of the dermohmmeter to an accuracy of 200 ohms and recorded for each trial: (R_b) the resistance at the time of the CS onset and (R_a) the lowest resistance attained following the CS onset until approximately 3 sec. after its termination.

RESULTS

For each *S* the two resistance measures recorded on each trial were used to obtain a log conductance (LC) value and a Conditioning Index (CI) score (Hunt, 1962) as defined below.

$$\text{LC value} = \log\left[\left(\frac{1}{R_a} - \frac{1}{R_b}\right)10^7 + 1\right]$$

and

$$\text{CI score} = \frac{\text{LC value}}{\text{mean UCR}},$$

where the mean UCR was based upon the LC values of the last eight S_+ acquisition trials.

An analysis of these last eight S_+ acquisition trials revealed that the effects of neither the ISI, $F\ (5, 98) = 1.21$, $p > .25$, nor the ISI × Trials interaction, $F < 1.0$, was significant, although the main effect of Trials was significant, $F\ (7, 336) = 4.74$, $p < .001$.

Estimates of the CR_+ and CR_- at two stages of differential conditioning were obtained by calculating the mean LC value and mean CI score on the test trial and Trial 12 (for the intermediate stage) and Trials 24 and 25 (for the terminal stage). These means are plotted as a function of the ISI in Fig. 2. An analysis of the CR_+ showed the effects of ISI to be significant, $F\ (5, 48) = 3.41$, $p < .05$, for the CI scores, but not for the LC values, $F\ (5, 48) = 2.12$, $p > .05$. Neither the effects of Conditioning Stage nor the Stage × ISI interaction was significant.

A similar analysis of the CR_- showed no significant effects involving either the ISI or the conditioning level. However, including all 12 S_- acquisition trials in the analysis of the CI scores produced significant main effects of both ISI, $F\ (5, 48) = 3.40$, $p < .05$, and Trials, $F\ (11, 528) = 2.52$, $p < .01$. A comparable significance of the effects of Trials, $F\ (11, 528) = 2.38$, $p < .01$, on the LC value was observed, but not for the effects of the ISI. For neither the LC values nor the CI scores was the ISI × Trials interaction significant.

Indexes of the DR at the two stages of conditioning were determined: (DR_1) the difference between the Test Trial CR_+ and the Trial 12 CR_- and (DR_2) the difference between the Trial 25 CR_+ and the Trial 24 CR_-. An analysis of variance revealed no significant difference between DR_1 and DR_2 ($F < 1$) for either the LC values or the CI scores. The combined means of DR_1 and DR_2 as a function of ISI are shown in Fig. 3. The analysis of the DR based on the CI scores revealed only the effects of ISI to be significant, $F\ (5, 48) = 2.70$, $p < .05$. For the

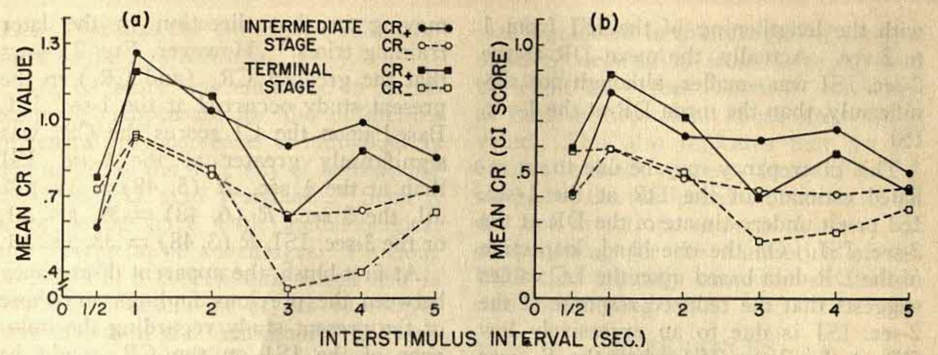


FIG. 2. Response to the S_+ and S_- based upon the (a) log conductance value and (b) conditioning index score, as a function of the interstimulus interval at two stages of differential conditioning.

LC values, only the interaction between the ISI and the tone employed as the S_+ was significant, $F\ (5, 48) = 2.42$, $p < .05$. The interaction appears to be attributable to a greatly reduced DR at the 2-sec. ISI when the S_+ was the 1,967-cps. tone; and the liberty of combining the data for both S_+ tones for plotting the function for the LC values in Fig. 3 has been taken.

Duncan's new multiple-range test (Steel & Torrie, 1960; Winer, 1962) revealed that the DR based on the CI scores at the $\frac{1}{2}$-sec. ISI was significantly different from both the 1-sec. ISI, $R\ (5, 48) = .250$, $p < .05$, and the 4-sec. ISI, $R\ (6, 48) = .278$, $p < .05$; while the DR at the 5-sec. ISI was different only from the 4-sec. ISI, $R\ (5, 48) = .220$, $p < .05$. The application of Duncan's test to the DR based on the LC values provides essentially the same results, with the exception that a separate analysis of the data for $S_+ = 1{,}967$-cps. tone revealed the only significant difference to be between the 2-sec. and 4-sec. ISI, $R\ (6, 48) = .444$, $p < .05$.

An analysis of the six extinction trials showed only the effects of Trials to be significant, $F\ (5, 240) > 7$, $p < .001$, for both the LC values and the CI scores; the interaction between Trials and ISI was not significant.

DISCUSSION

The results are generally consistent with the hypothesis that a maximum DR occurs at some ISI longer than the approximately $\frac{1}{2}$-sec. interval typically reported for simple conditioning and that the DR decreases with further extensions of the ISI. In the present study, the DR at both the shortest ISI ($\frac{1}{2}$ sec.) and the longest ISI (5 sec.) was significantly smaller than at the 4-sec. ISI. Hartman and Grant (1962) and Kimmel and Pennypacker (1963) provide data for the 2-sec. ISI and shorter with which the present results may be compared. There is a general compatibility in that the $\frac{1}{2}$-sec. ISI exhibited a smaller DR than either the 1- or 2-sec. intervals. However, it was expected on the basis of the finding of Kimmel and Pennypacker that an increased DR would be associated

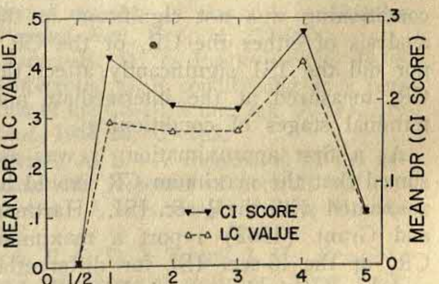


FIG. 3. Mean differential GSR to the S_+ and S_- based upon the log conductance values and the conditioning index scores, as a function of the interstimulus interval.

with the lengthening of the ISI from 1 to 2 sec. Actually, the mean DR at the 2-sec. ISI was smaller, although not significantly, than the mean DR at the 1-sec. ISI.

This discrepancy may be due to an inflated estimate of the DR at the 1-sec. ISI or an underestimate of the DR at the 2-sec. ISI. On the one hand, inspection of the DR data based upon the LC values suggests that the reduced response at the 2-sec. ISI is due to an excessively low DR at the 2-sec. ISI when the S_+ was the 1,967-cps tone. On the other hand, the relatively small magnitudes of the DR at the 3- and 4-sec. ISI lend credibility to the notion that the DR at the 1-sec. ISI is overestimated. The data of the present investigation do not provide a basis for choosing between the two possibilities.

The DR was not significantly different at the two stages of differential conditioning. However, inspection of Fig. 2 suggests that although the differences between the CR_+ and CR_- functions at the various ISIs may be approximately the same at both stages of conditioning, there is an overall depression of both the CR_+ and the CR_- at the 3-, 4-, and 5-sec. intervals associated with increased differential conditioning. Also, it appears that with increased conditioning the CR_- may attain some minimum at, say, 3 or 4 sec. and then increase with further extensions of the ISI. However, statistical support for these observations is lacking since the interaction between ISI and stages of conditioning was not significant in the analysis of either the CR_+ or the CR_-; nor did the ISI significantly affect the CR_- measured at the intermediate and terminal stages of conditioning.

As a first approximation, it was assumed that the maximum CR_+ would be associated with the ½-sec. ISI. Hartman and Grant (1962) report a maximum CR_+ at the .6-sec. ISI for differential eyelid conditioning; and Kimmel and Pennypacker (1963) indicate, "the conventional CS-UCS interval function for the positive CS was not as readily discernible, although the data seemed to be moving in that direction in the later training trials." However, Fig. 2 shows that the greatest CR_+ (and CR_-) in the present study occurred at the 1-sec. ISI. Based upon the CI scores the CR_+ was significantly greater at the 1-sec. ISI than at the ½ sec., $R\ (5, 48) = .34$, $p < .01$, the 5 sec., $R\ (6, 48) = .39$, $p < .01$, or the 3-sec. ISI, $R\ (3, 48) = .33$, $p < .01$.

At first blush, the apparent discrepancy between the previous findings and those of the present study regarding the influence of the ISI on the CR_+ might be attributed to the relative differences between the S_+ and S_- values employed. For example, Kimmel and Penneypacker (1963) used 800- and 1,000-cps. tones and the present study employed 1,000- and 1,967-cps. tones. This would suggest that there is some interactive effect of the ISI and the S_-, e.g., an inhibitory tendency associated with the S_-, on the CR_+; and thus the ISI at which the maximum CR_+ would occur depends upon the similarity of the S_+ and S_-. The present study which used more widely different S_+ and S_- values should result in a reduced influence of the S_- on the CR_+; thus, the conventional single stimulus ISI function for the CR_+ should emerge at least as clearly, and perhaps more so, in the present study than in that of Kimmel and Pennypacker (1963). Although the slight decrease in peakedness at the 1-sec. ISI associated with the terminal stage (Fig. 2) is consistent with this expectation, the ISI × Conditioning Stage interaction was not significant.

A comparison of the present findings with those of previous investigations is further complicated by differences in the introduction of nonreinforced test trials for the S_+ during acquisition. Considering only the results on the first test trial, which avoids any contamination due to nonreinforced S_+ trials, the findings of the present study and those of Kimmel and Pennypacker (1963) are quite similar. This suggests that an acceptable concept of the differential conditioning process needs to incorporate more explicitly not only the S_+/S_- ratio (Gynther, 1957), but also the percentage reinforcement of the S_+ (Newman, 1966).[2]

The CI score was employed in an attempt to compensate for the differences in general responsiveness of individual *S*s and minimize the necessity of considering the base-level skin resistance. Such a score would have both methodological and interpretative advantages. Previous proposals of a conditioning index that is relatively independent of *S*'s base level have employed the "maximum" GSR as the value against which to compare other responses (Elliot & Singer, 1953; Paintal, 1951); the "maximum" was elicited by either a single faradic shock or loud noise. The rationale of the present CI score was to obtain a more reliable index by calculating the mean of a number of responses to the shock. A crucial assumption underlying the use of the CI score is that the mean UCR is not systematically related to the independent variable(s) being studied.

Kimble and Ost (1961) report that the *gain* in the UCR amplitude on UCS-alone trials following eyelid conditioning trials at ¼-, ½-, 1-, and 2-sec. ISIs reflects the conventional ISI conditioning function. Similarly for GSR conditioning, Kimmel (1965, 1966) found that the maximum UCR-recovery occurs following training ½-sec. ISI, although only marginal statistical significance ($p < .10$) is reported. This suggests that the mean UCR measure employed in the present study as the denominator of the CI score may be related to the ISI and thus fail to satisfy the crucial assumption indicated above. However, in the present study, the LC values upon which the mean UCR was based were not significantly affected by either the ISI or the ISI × Trial interaction. The discrepancy between the findings of Kimble and Ost (1961) and Kimmel (1965, 1966), on the one hand, and the present study, on the other, cannot be accounted for in a simple fashion.

The fact that the main effect of trials was significant in the analysis of the last eight S_+ acquisition trials indicates that the eight LC values cannot be conceived as random dispersions about some single value. This also indicates that the increases in shock current (from 1.7 to 2.5 ma.) over the reinforced acquisition trials did not attain its desired objective of compensating for the diminution of the UCR that typically occurs over successive conditioning trials.

REFERENCES

ELLIOT, D. N., & SINGER, E. G. The Paintal Index as an indicator of skin-resistance changes to emotional stimuli. *J. exp. Psychol.*, 1953, **45**, 429–430.

GYNTHER, M. D. Differential eyelid conditioning as a function of stimulus similarity and strength of response to the CS. *J. exp. Psychol.*, 1957, **53**, 408–416.

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

HUNT, D. P. The effect of stimulus similarity on the acquisition and extinction of a conditioned response. *J. exp. Psychol.*, 1962, **63**, 278–283.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KIMBLE, G. A., & OST, J. W. A conditioned inhibitory process in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 150–156.

KIMMEL, H. D. Instrumental inhibitory factors in classical conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965. Pp. 148–171.

KIMMEL, H. D. Inhibition of the unconditioned response in classical conditioning. *Psychol. Rev.*, 1966, **73**, 232–240.

KIMMEL, H. D., & PENNYPACKER, H. S. Differential GSR conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1963, **65**, 559–563.

PAINTAL, A. S. A comparison of the galvanic skin responses of normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.

STEEL, R. G. D., & TORRIE, J. H. *Principles and procedures of statistics.* New York: McGraw-Hill, 1960.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 6, 1966)

---

[2] Differential eyelid conditioning as a function of the probability of reinforcement, unpublished manuscript, 1966.

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 26–30

EFFECT OF DURATION OF VIEWING ON FORM AND SIZE JUDGMENTS[1]

RUDY NESMITH AND ALBERT S. RODWAN[2]

Emory University

2 microgenetic hypotheses were tested: perception would become more stable with an increase in duration of viewing; and this increase is different for form and for size. 100 plane figures, all combinations of 10 heights and 10 widths varying in increments of .01 in., constituted the stimulus set. They were presented singly as back-lighted figures in a completely dark surround. 7 male Ss had at least 20-20 vision uncorrected or with contact lens. There were 9 durations of viewing and 2 types of judgments: form, squares and rectangles, or size, large and small. Discrimination measures were computed and showed that Ss could discriminate form better than they could discriminate size. There was no significant effect of duration on either type of discrimination, nor was there any interaction between type of judgment and duration.

Studies in the microgenesis of perception have been oriented toward analyzing the events or stages occurring within the temporal period between the presentation of the stimulus and the formation of a stable perception. Heinz Werner (1956) has made the basic assumption of microgenesis explicit: Perception is a time dependent process and can be partitioned into separate sequential stages of development. Furthermore, if the duration of viewing is too short no stable perception can occur.

Sander (1928, 1930) developed the microgenetic approach and conducted several experiments which he used as evidence for a time dependence in perception. Sander suggested that the perceptual process undergoes four major stages of development as a function of the duration of viewing. The initial stage is a "diffuse undifferentiated whole." The second stage involves a figure-ground separation; this is followed by an increasing distinctiveness of contour and "inner content." Finally, with further elaboration on the inner content, the perception is complete (Sander, 1930; Undeutsch, 1942).

Historically, the experimental paradigm for migrogenetic studies has been to present complex stimuli such as paintings, geometric forms, or common objects at different durations of viewing. The typical responses were introspective description, drawings or other reproductions, or manual manipulations of the objects themselves. Flavell and Draguns (1957) reviewed a great deal of the research in this area and report it as being "quite poor by present-day methodological standards."

Various extensions of the microgenetic approach have been offered. In 1955 Kragh related microgenetic development to ontogenetic development. He suggested the adult's perception of an object goes through the same stages as does the child's perception of the same object. The perception that an adult has of a briefly presented stimulus is the same as the perception that a child would have when he viewed the stimulus for a much longer period of time. The logic is that the mature perceptual system would show a "regression"-like effect when duration of viewing is restricted.

---

[1] An earlier version of this paper was read at the SEPA meeting in New Orleans in March 1966.

[2] This research was supported by funds from National Institutes of Health Grant No. MH 11210.

At the bottom of all microgenetic hypotheses is the assumption that perceptual ability, veridicality, sensitivity, etc., increases as a function of the duration of viewing. Another hypothesis which derives from microgenetic-ontogenetic parallelism is that the microgenetic development of form perception would be different than for size perception. This follows from the assumption that form is more basic than size; that is to say, form occurs earlier in ontogenetic development than does size. Size judgments, since they involve abstract notions of area and other logical operations, would develop later and, due to the parallelism, would decrease as the duration of viewing increased.

The following study is an attempt to systematically investigate these two hypotheses: Does discriminating increase as a function of the duration of viewing? And, is the form-time function different from the size-time function?

METHOD

Stimuli.—The stimulus set consisted of 100 plane figures. There were 10 heights and 10 widths which were paired in all possible combinations. The figures were squares and rectangles which varied from 2.16 in. to 2.26 in. in increments of .01 in.

Apparatus.—Using the method of absolute judgment, the stimuli were presented to *S* at a distance of 36 in. in the middle arm of a three arm tachistoscope. The middle arm has a rectangle generator built into it. The center of the generator is fixed and the figure is varied by moving the edges which are geared doors that close toward the center. By setting both the vertical and horizontal doors, *E* is able to select any stimulus from the set for presentation. The maximum error in each setting is less than .03 mm. The figure is backlighted and appears as a luminous figure in a black surround. The light source is from two 4-w. fluorescent bulbs.

The *S* places his head in a viewing mask which is mounted on the tachistoscope; this mask stabilizes the *S*s head and occludes the ambient room light, and occludes *S*s left eye. Before each trial *S* fixates a dot of light which is located where the center of the stimulus will appear. The fixation point goes off 1 sec. before the stimulus figure comes on. During that second, the field is completely dark; similarly, after the offset of the stimulus the field is completely dark. Thus the fixation point serves two purposes: to orient *S* and to serve as a warning signal.

Subjects.—The *S*s were seven male college students. None had served in an experiment before. Each *S* was checked for visual acuity in the right eye, uncorrected, or with contact lens. The visual acuity check occurred prior to the first experimental session.

Procedure.—Each *S* went through 18 sessions. This consisted of all combinations of nine durations of viewing with two classes of judgments. The durations of viewing were: .01, .02, .04, .08, .16, .32, .64, 1.28, and 2.56 sec. The two classes of judgments were form (deciding whether the figure was a square or a rectangle) and size (deciding whether the figure was large or small). The criteria for these judgments were never made explicit; each *S* was given the responses which were appropriate for that session but was never given any feedback.

The order of presentation of the sessions was randomized and the order of presentation of each stimulus within each session was also randomized, independently. This was done for each *S*. At the beginning of each session *S* was given 20 practice trials; the stimuli which were used in the practice trials were presented as the last 20 in the experimental trials. This gave a different set of practice stimuli for each session.

The intertrial interval was approximately 8 sec. In this period *S* looked at the figure for the appropriate duration, *withdrew his head when the figure went off*, then he immediately wrote his judgment on his own response sheet and responded orally to *E*. Finally, when he had completed his responding, he positioned himself (faced into the tachistoscope) for the next trial. The next trial was begun when *S* reported that he was ready. The experimental room was painted flat green and illuminated with four 100-w. incandescent bulbs. The dimensions of the experimental room were: 10 ft. wide, 15 ft. long, and 10 ft. high.

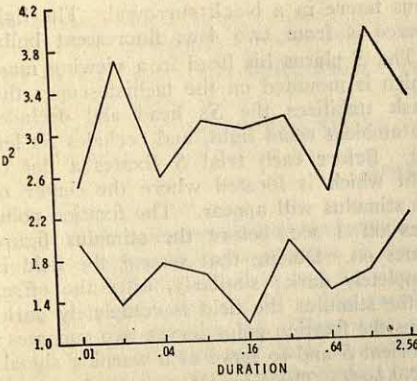


FIG. 1. Average D^2 plotted against duration of viewing. (The upper curve is for form and the lower curve is for size.)

Two sessions were run per day for each of the *S*s. These were separated by at least a 2-hr. period in which other *S*s were run.

Analysis.—A D^2 statistic was computed for each *S* on each session. This statistic is a measure of the distance separating the means of two multivariate categories on a weighted linear decision axis (Rodwan & Hake, 1964). D^2 is a ratio of the variance between means to the variance within classes. Thus it is a standardized measure and uses the within class variance as a unit of measure (metric). D^2 can be thought of as a measure of confusion: the larger D^2 is, the less the confusion; similarly, it can be considered as a multivariate analogue of the familiar *d′* statistic (Swets, 1964). The difference between these two statistics is that *d′* is a standardized measure of the contingency between *S*s' responses and *E*'s criterion. On the other hand, D^2 is a standardized measure of *S*s confusion in satisfying his own criterion. Thus, the distance between two means on an experimentally defined axis is given by *d′* whereas D^2 gives the distance between means on *S*s own decision axis.

RESULTS

Figure 1 is a graph of the size-time and the form-time functions. It can be seen that the D^2s for form are larger than for size for each duration of viewing. The probability of this happening by chance (using the sign test) is .002. There was less confusion in the form judgments than in the size judgments; on the average, *S*s could separate squares from rectangles better than they could separate large from small figures. A more detailed analysis was performed. Each *S*s data were analyzed using matched-pairs *t* test. This compared the D^2 for form with the D^2 for size for each duration of viewing. Of the seven *S*s, five showed a differential ability to separate form better than size ($p < .05$). The other two *S*s did not show any differential ability.

The second analysis was a three-way factorial design mixed model with rows being judgments, columns being durations and slices being *S*s. Each *S* had all 18 conditions thus making this identical to the design given by Lindquist (1956) on page 237. Both Lindquist (1956) and Edwards (1964) agree on the appropriate error terms: for the judgments, the error term is the *S*s × Judgment interaction; for the duration effects, the error term is the *S*s × Duration interaction; for the Judgments × Duration interaction effect, the error term is the three-way interaction (Edwards, 1964, p. 100).

This analysis is given in Table 1. It can be seen that the judgments are significantly different. This corroborates the sign test done on the means in Fig. 1. It is interesting to note that Dura-

TABLE 1

A THREE FACTOR MIXED MODEL ANALYSIS OF VARIANCE: JUDGMENTS AND DURATION OF VIEWING ARE FIXED AND *S*s ARE RANDOM

Source	*df*	*MS*	*F*
Judgments (J)	1	23.28	9.54*
Durations (T)	8	.77	1.64
Subjects (S)	6	4.22	30.14*
J × T	8	.39	2.79
J × S	6	2.44	17.45*
T × S	48	.47	3.36
J × T × S	48	.14	
Total	125		

* $p < .05$.

TABLE 2

Two Treatment × *S*s Analyses of Variance

Source	*df*	*MS*	*F*
Form			
*S*s	6	4.4962	10.93**
Duration	8	.8630	2.10
Residual	48	.4114	
Size			
*S*s	68	2.0924	10.86**
Duration	8	.3054	1.58
Residual	48	.1927	

** $p < .01$.

tion of Viewing and Judgment × Duration interaction are not significant. The implication is that the two curves in Fig. 1 are parallel to one another and to the time axis.

There is one point which needs further consideration. It is possible that a slight, but significant, interaction exists for each *S* but washes out in the averaging for error term. To check this out, two Treatment × *S*s analyses were performed on these data, one for form and one for size. This is presented in Table 2. It can be seen there that the only significant source of variation was the *S*s. Thus the Treatments × *S*s design confirms the result: duration of viewing has no effect on the magnitude of D^2.

Discussion

These results fail to suport the microgenetic hypothesis concerning the effect of duration of viewing. They also fail to support the hypothesis of interaction between types of judgment and duration of viewing. The only positive finding is that judgments concerning the form of a set of stimuli involve less confusion than size judgments of the same set. These data suggest that since D^2 for the square-rectangle judgments is larger than for the large-small judgments, the stimuli can be organized into form categories easier, i.e., that form categories are more stable, than size.

These results confirm and extend results from previous studies in this laboratory. In the first study Rodwan (1964) showed that when *S*s were required to assign plane figures into one of four categories. Large-square, large-rectangle, small-square, and small-rectangle, that the form effect produced a larger *F* ratio than size. This was true for 8 of 10 *S*s. He interpreted this to mean that *S*s have a greater ability to discriminate form. In the second study Rodwan (1965) *S*s were required to classify plane figures into one or two categories. The *S* was not given any instructions as criteria. He was simply told to group the stimuli into two categories such that within categories, there would be minimum difference and between categories maximum difference. They were to use "X" and "Y" as response labels. He found that all nine *S*s used a form criterion, i.e., their judgments correlated .971 with a form decision axis.

It is possible that a duration of .01 sec. is too long a time interval for the relatively simple figures used in this study. The typical stimulus figure used in microgenetic studies is a complex one; one of the favorites is to use photographs of known objects. It may be that complex figures follow a different law with respect to development. Another difference between this experiment and the typical microgenetic ones is in the use of a fixation point. It may be that the combination of simple figure and fixation point produced a too rapid development of a stable perception. On the other hand, it is possible that the microgenetic hypothesis can be subsumed under other, well known, principles. For example, it is clear that when the signal is not explicitly located in his visual field, *S* requires a certain amount of time to search and scan his field, in order to find the signal. It could be that the microgenetic effects are due to this search and scan procedure.

REFERENCES

Edwards, A. L. *Expected values of discrete random variables and elementary statistics.* New York: Wiley, 1964.

FLAVELL, J. H., & DRAGUNS, J. A. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

KRAGH, U. *The actual-genetic model of perception-personality.* Lund: CWK Leerup, 1955.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

RODWAN, A. S. An empirical validation of the concept of coherence. *J. exp. Psychol.*, 1964, **68**, 167–170.

RODWAN, A. S. Primacy of form criterion in perceptual judgments. *J. exp. Psychol.*, 1965, **70**, 231–232.

RODWAN, A. S., & HAKE, H. W. The discriminant function as a model in perception. *Amer. J. Psychol.*, 1964, **77**, 380–392.

SANDER, F. Experimentelle Ergebnisse der Gestaltpsychologie. *Ber. ü.d. Kongr. f. exper. Psychol.*, 1928, **10**, 23–87.

SANDER, F. Structures, totally of experience, and gestalt. In C. Murchisun (Ed.), *Psychologies of 1930,* Worcester, Mass.: Clark Univer. Press. 1930. Pp. 188–204.

SWETS, J. A. *Signal deduction and recognition by human observers.* New York: Wiley, 1964.

UNDEUTSCH, U. Die aktualgenese in ihrer allgemeinpsychologishen und ihrer charakterologischen bedeutung. *Scientia,* 1942, **72**, 37–42, 95–98.

WERNER, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, **52**, 347–353.

(Received May 6, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 31–35

UCS INTENSITY AND THE LATENCY OF THE CLASSICALLY CONDITIONED EYELID RESPONSE[1]

MILTON D. SUBOSKI[2]

Indiana University

Human Ss were classically conditioned at 3 air-puff intensities with ½ the Ss detecting a signal embedded in the CS. The higher UCS intensities produced higher frequency and shorter latency CRs and detection served to inhibit responding and increase latency. CR latency decreased during acquisition and increased during extinction; results interpreted as indicating that the CR develops near UCS onset, moves toward CS onset as it is strengthened, and then moves back toward UCS onset as it is weakened.

This report is primarily concerned with the effects of UCS intensity and acquisition and extinction training on CR latency. In the only study reporting UCS intensity effects on latency, Gormezano and Moore (1962) found decreasing latency with increasing UCS intensity. No clear picture of the effects of training on CR latency emerges from the experimental literature. Early results (Hilgard & Campbell, 1936; Hilgard & Marquis, 1935; Pavlov, 1927) indicated that CR latency decreases in acquisition and increases in extinction. Later work has produced mixed results for acquisition training, with both decreasing (Ebel & Prokasy, 1963; Schneidermann & Gormezano, 1964) and increasing (Boneau, 1958; Runquist & Muir, 1965; Runquist & Towart, 1965) latency reported. Spence (1956) has indicated that no consistent changes in latency had been observed in a large number of studies coming from his laboratory, and further that latency changes seemed to be related to CR frequency with high-level conditioners showing decreased latency with trials and low-level conditioners increased latency.

The present experiment compares human eyelid conditioning performance by *S*s detecting a signal embedded in the CS with performance of nondetecting *S*s at three UCS intensities. The detection task was used with reference to Broadbent's (1953) suggestion that classical conditioning phenomena could be explained in terms of vigilance effects and also to the hypothesis that UCS intensity affects response probability by increasing the number of trials on which *S* attends to the CS. Since the results were essentially negative with respect to these hypotheses, no further consideration is given to them (see Suboski, 1966, for further details).

METHOD

Subjects.—The *S*s were 60 male and 60 female volunteers from an introductory laboratory course in psychology, participating for extra credit.

Apparatus.—The *S* sat in a straight-backed wooden chair in one of two identical 5-ft. by 8-ft. rooms, facing at a distance of approximately 3½ ft., a 4-in. diameter disk

---

[1] This report is based on a dissertation submitted to the Graduate School of Indiana University in partial fulfilment of the requirements of the PhD degree. I am grateful to I. Gormezano who directed the research and provided support from National Science Foundation Grant GB-2843. Portions of these data were presented at the Midwestern Psychological Association Meeting, May, 1966.

[2] Now at Queen's University, Kingston, Ontario.

of $\frac{1}{4}$-in. white opaque plastic, mounted flush in a flat-white plywood surface at eye level. Four General Electric NE 45 neon bulbs were located $\frac{1}{4}$ in. behind the disk, forming a $1\frac{3}{4}$-in. sided square, centered on the center of the disk, with the bulbs located on the vertical and horizontal axes.

The *S* wore an elastic headband which supported an airjet with a 2-mm. diameter orifice and a minitorque potentiometer. The airjet was pointed directly at the surface of the cornea from a position slightly below and approximately $\frac{5}{8}$ in. distance. Flexible coupling of the barrel of a No. 19 syringe needle to the shaft of the potentiometer permitted a piece of light wire, loose fitted into the barrel of the needle, to be attached to *S*'s right upper eyelid by means of a small piece of tape. The signal from the potentiometer produced by movement of *S*'s right eyelid was amplified and recorded by an ink-writing Brush penmotor at a paper speed of 200 mm/sec, where CS and UCS onset and offset were also recorded. The *S* also wore a set of Grason-Stadler D30 earphones, through which white masking noise of approximately 40-db. SPL was maintained continuously for the duration of the session.

Stimuli durations were controlled by Hunter timers and relay circuitry, and intertrial intervals and pattern of delayed lights by Western Union tape transmitters. Individual *S* microphones, and a stereophonic headset for *E* permitted scoring of *S*'s verbal responses.

Procedure.—Twenty *S*s were randomly assigned to each of six groups with the restrictions that five males and five females within each group were run in each of the two *S* rooms.

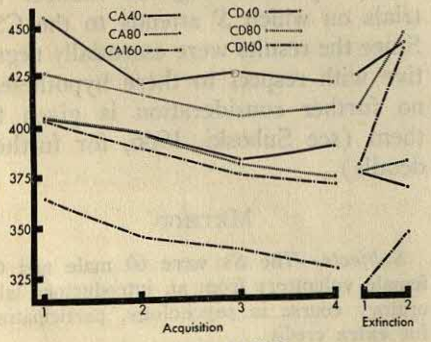


FIG. 1. Mean response latency in milliseconds from CS onset for all groups, plotted in 20-trial blocks in acquisition and 10-trial blocks in extinction.

On each CS presentation, three of the lights came on, followed after a 67-msec. delay by the fourth. The total presentation time was 575 msec., with the last 100 msec. overlapping with the air puff. The delayed light was randomly preselected with the restrictions that in the 100 trials each light was delayed 25 times, and no light was delayed more than four times in succession. Three groups (Groups CD40, CD80, CD160) received 100 CS presentations on which they were instructed to detect and report which light was delayed and also received UCS presentations paired with the first 80 CS presentations, followed by 20 CS alone extinction trials. The UCS was a puff of compressed nitrogen at an intensity, measured at the point of delivery to the eye, sufficient to produce a maximum deflection in a column of mercury of 40, 80, or 160 mm. for Groups CD40, CD 80, and CD160, respectively (.77, 1.55, 3.09 psi, respectively). Three groups (Groups CA40, CA80, CA160) received identical stimuli presentations but were not required to perform the detection task, nor was the delayed light mentioned to them. The intertrial interval for all groups was 15, 20, or 25 sec., randomized in blocks of three.

Standard neutral eyelid-conditioning instructions (Gormezano & Moore, 1962) and special instructions were read to *S*s over an intercom just prior to the start of the session. The latencies of eyelid closures producing recording pen deflections of at least 1 mm. from the base line were scored, and these in the latency range of 150 to 500 msec. were considered to be CRs.

RESULTS

Analysis of variance of the arc sine transform of individual percentages of acquisition CRs revealed UCS Intensity, *F* (2, 96) = 10.87, Trials, *F* (3, 288) = 146.09, and Trials × Detection Condition interaction, *F* (3, 288) = 6.22 (all p's $< .01$) effects. A similar analysis of extinction data showed the Detection Condition, *F* (1, 96) = 5.66, $p < .05$, and UCS Intensity, *F* (2, 96) = 5.82, $p < .01$, main effects to be significant. Trial blocks were also a highly significant source of variation, but not the Trials × Detection Condition interaction. These effects all reflected the higher levels of responding produced by the higher UCS intensities, and the generally lower levels of responding by *S*s performing the detection task.

Figure 1 presents mean CR latency for CA and CD groups plotted in 20 trial blocks in acquisition and 10 trial blocks in extinction. The increased CR latency produced by the detection task is shown quite clearly, although the decrease in latency across trial blocks in acquisition appears comparable for both CA and CD groups. The results of an unweighted means, unequal N analysis of variance performed on the logarithm of the individual mean latencies in acquisition indicated that the Detection Condition, $F\ (1,\ 92) = 23.68$, UCS Intensity, $F\ (2, 92) = 15.77$, and Trial Blocks, $F\ (3,\ 27) = 24.54$, p's $< .01$, all produced significant effects, reflecting the shorter latencies of the CA groups, the higher UCS intensities, and the later trial blocks.

Figure 1 shows a striking increase in latency across trial blocks by the CD groups during extinction. Analysis of trial effects is complicated, however, by the fact that only a small number of *S*s gave CRs in the second trial block of extinction. Of the 60 Group CD *S*s, only 22 gave CRs in both trial blocks; 21 of whom had shorter mean latencies in the first trial block, a result which is significant beyond the .01 level by a sign test. Not obvious from Fig. 1 is the fact that of the 30 CA *S*s who gave CRs in both trial blocks, 24 showed longer latencies in the second, also significant beyond the .01 level by a sign test. Detailed analysis indicates that initial latency of CRs in extinction is related to resistance to extinction. The *S*s showing long initial latencies tend to stop giving CRs rather quickly, and *S*s showing shorter latencies tend to continue giving CRs but with progressively longer latencies. For CA *S*s, the mean latency for *S*s giving CRs only in the first trial block is 379.5 msec. and for *S*s giving CRs in both trial blocks, 335.6 msec. The same effect occurred in the CD *S*s, and in fact, for each of the groups separately. The combination of *S*s showing initial long latencies dropping out and *S*s who continue to give CRs showing longer latencies apparently produces the small shifts in latency between trial blocks shown by the CA groups in Fig. 1.

In view of the relationship between latency and resistance to extinction, correlations between CR latency and frequency were obtained for the last 20 trials of acquisition for the CA groups. The values were: Group CA40, $r = -.73\ (p < .01)$; Group CA80, $r = -.76\ (p < .01)$; and Group CA160, $r = -.28$.

Spence (1956, pp. 70–71) presented latency curves for *S*s selected on the basis of response frequency which show decreasing latency across trials for high-level conditioners and increasing latency for middle and low-level conditioners. Figure 2 presents similar curves for each of the CA groups. The two *S*s showing the lowest frequency of CRs were eliminated, and means obtained by dividing the remainder into successive thirds on the basis of CR frequency. Figure 2 shows reasonably clearly that mean latency is inversely related to CR frequency and that latency is a decreasing function of trials.

DISCUSSION

The decreasing CR latency with increasing UCS intensity supports the findings of Gormezano and Moore (1962). The decreasing CR latency with increasing amounts of training has been previously shown by Hilgard and Campbell (1936), for the rabbit nictitating membrane (Schneidermann & Gormezano, 1964), for the dog eyelid (Hilgard & Marquis, 1935), and for extended periods

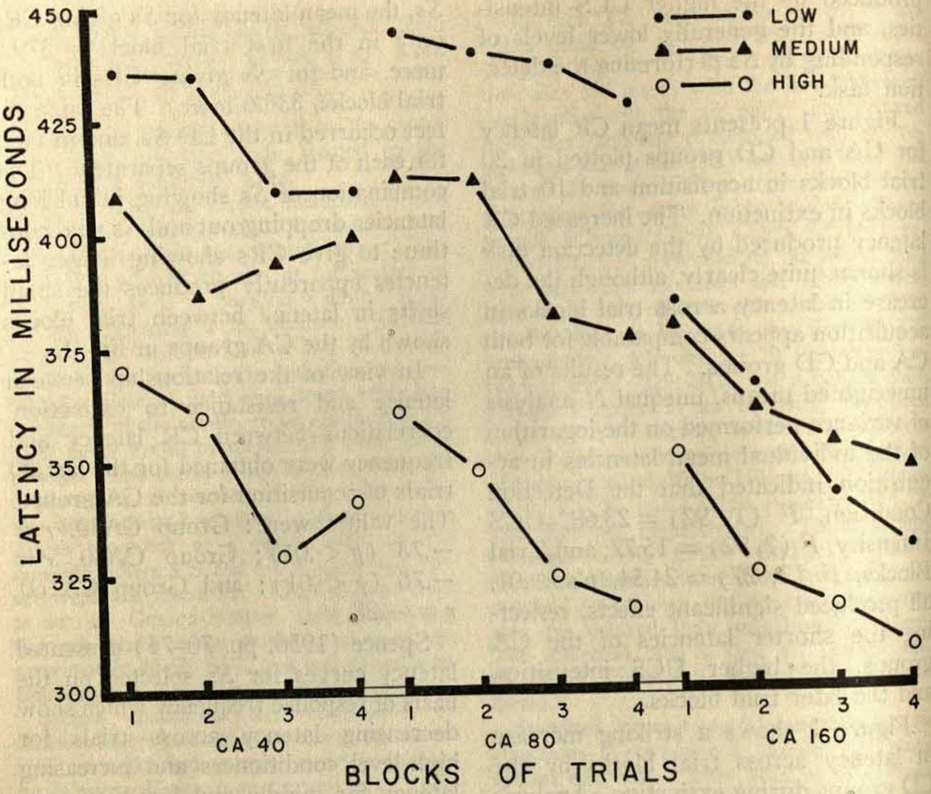


FIG. 2. Mean response latency for CA *S*s divided into successive thirds on the basis of CR frequency.

of human eyelid conditiong (Ebel & Prokasy, 1963), and contrasts with the results reported by Boneau (1958) for human eyelid conditioning. The inverse relationship between CR latency and frequency has been previously reported for human eyelid conditioning (Campbell & Hilgard, 1936; Pennypacker, 1964) and for the rabbit nictitating membrane (Schneidermann & Gormezano, 1964). The findings of increasing CR latency during the course of extinction is relatively unusual. Pavlov (1927) found such an effect in salivary conditioning, Hilgard and Campbell (1936) for the human eyelid, Hilgard and Marquis (1935) for the dog eyelid, and Prokasy, Fawcett, and Hall (1962) have recently reported increasing latency in extinction following GSR conditioning. Together these findings suggest that in the simple classical conditioning situation the CR develops near the point of UCS onset, moves in the direction of CS onset as it is strengthened, and then moves back in the direction of UCS onset as it is weakened.

REFERENCES

BONEAU, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

BROADBENT, D. E. Classical conditioning and human watchkeeping. *Psychol. Rev.*, 1953, **60**, 331–339.

CAMPBELL, A. A., & HILGARD, E. R. Individual differences in ease of conditioning. *J. exp. Psychol.*, 1936, **19**, 561–571.

EBEL, H. C., & PROKASY, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *J. exp. Psychol.*, 1963, **65**, 52–58.

Gormezano, I., & Moore, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, **63**, 487–494.

Hilgard, E. R., & Campbell, A. A. The course of acquisition and retention of conditioned eyelid responses in man. *J. exp. Psychol.*, 1936, **19**, 227–247.

Hilgard, E. R., & Marquis, D. G. Acquisition, extinction and retention of conditioned lid responses to light in dogs. *J. comp. Psychol.*, 1935, **19**, 29–58.

Pavlov, I. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

Pennypacker, H. S. Measurement of the conditioned eyelid reflex. *Science*, 1964, **144**, 1248–1249.

Prokasy, W. F., Fawcett, J. T., & Hall, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of interstimulus interval. *J. exp. Psychol.*, 1962, **64**, 513–518.

Runquist, W. N., & Muir, W. R. Intrasession decrements in the performance of the classically conditioned eyelid reflex. *J. exp. Psychol.*, 1965, **70**, 520–525.

Runquist, W. N., & Towart, E. Further evidence for intrasession decrements in eyelid conditioning. *Psychol. Rep.*, 1965, **17**, 891–897.

Schneiderman, N., & Gormezano, I. Conditioning of the nictitating membrane of the rabbit as a function of CS-US interval. *J. comp. physiol. Psychol.*, 1964, **57**, 188–195.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Suboski, M. D. An investigation of stimulus selection processes in human eyelid conditioning. Unpublished doctoral dissertation, Indiana University, 1966.

(Received May 7, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 36–43

INTERACTION BETWEEN INSTRUCTIONS AND ISI IN CONDITIONING AND EXTINCTION OF THE GSR[1]

IRWIN J. MANDEL AND WAGNER H. BRIDGER

Department of Psychiatry, Albert Einstein College of Medicine, Yeshiva University

90 human Ss acquired a differentially conditioned GSR with an instructed acquisition procedure under 3 different ISIs: backward (UCS-CS), 500 msec. and 5 sec. forward. At the onset of extinction each ISI group was divided into Informed and Noninformed halves. A postexperimental inquiry eliminated those Informed Ss who did not believe the extinction instructions. The backward conditioned GSR did not demonstrate resistance to extinction under either extinction condition. Among the 2 forward ISI groups during extinction, the 5-sec. group demonstrated a larger differential response under the Noninformed condition while the 500-msec. group had a larger differential response under the Informed condition. Both forward ISI groups demonstrated differential extinction responses under the Informed condition contrary to their cognitive expectancy. Alternative theories are presented to account for the results.

The use of different interstimulus intervals (ISIs) in the classical conditioning of responses in human *S*s has produced a considerable amount of controversy. For the conditioned GSR, adequate acquisition performance has been reported using ISIs ranging from a backward (UCS-CS) interval of −2,800 msec. (Champion, 1962) to a forward interval of 20 sec. (Switzer, 1934). Kimmel and Pennypacker (1963) have reported better *differential* GSR conditioning with longer ISIs as have Hartman and Grant (1962) in eyelid conditioning.

In terms of resistance to extinction of the conditioned GSR, White and Schlosberg (1952) and Moeller (1954) have rigorously upheld the 400–500 msec. forward ISI as the only interval producing resistance to extinction. Wickens, Gehman, and Sullivan (1959) indicated two maxima in the forward ISI function, 500 msec. and 2,000 msec., while Prokasy, Fawcett, and Hall (1962) reported no difference in resistance to extinction for groups conditioned with forward ISIs from 500 msec. to 5 sec.; only a 0-msec. group failed to show any resistance. Champion and Jones (1962) compared resistance to extinction between a backward 750-msec. ISI group and a forward 500-msec. group and failed to find significant differences.

Although an *S* cannot exercise direct voluntary control over his GSR, Cook and Harris (1937) demonstrated that verbal control of the GSR is possible reporting that a GSR produced by threat of electric shock showed response strength equal to that of a GSR produced by 30 CS-UCS pairings. These investigators also reported that informing *S* as to the onset of extinction virtually eliminated resistance to extinction. Almost simultaneously, Mowrer (1938) reported a similar phenomenon; these two results led Hilgard and Marquis (1940, p. 206) to characterize the GSR as a unitary response due to a single underlying process, an expectancy state. Mowrer, however, suggested that it may be possible to establish a conditioned GSR which would not disappear as soon as the expectancy of further shock was not present. This author maintained that if this result could be obtained, a different form of learning ("true learning") would be demonstrated that is not based merely on a change in preparedness. Mowrer also briefly noted what has proven to be a critical experimental variable; the *S*'s belief in the information as to no further reinforcements.

---

[1] This article is based on a dissertation by the senior author submitted to the Ferkauf Graduate School of Education, Yeshiva University, in partial fulfillment of the requirements for the PhD degree. The research was supported by United States Public Health Service Grants MH-06282, K3-MH-4177, and 5TI-MH-6418. The authors gratefully acknowledge the assistance of the Section of Biomathematics, Department of Physiology in the use of a computer for both data reduction and analysis. The computer facility is supported by United States Public Health Service Grant NB-03491 and New York City Health Research Council Grant U-1077. A portion of this research was presented at the Eastern Psychological Association meetings, April 1965.

In a previous report, the present authors (Bridger & Mandel, 1965) presented a procedure for assessing *S*'s confidence in extinction information. Although dealing primarily with the partial-reinforcement effect, this experiment also suggested the existence of two levels of GSR conditioning that obey different psychological laws. If the two levels of GSR conditioning do exist (i.e., responses in accord with and responses contrary to a cognitive expectancy), then they may well interact with the ISI parameter and account for the conflicting reports about the effects of the ISI in GSR conditioning.

METHOD

Subjects.—The *S*s were 144 paid volunteer college students 18–25 yr. from various schools in the New York City area. Of this group 6 were eliminated due to experimental error, 13 refused to continue due to the intense shock, and the protocols of 35 others were excluded on the basis of the postexperimental questionnaire to be described in a later section. Of the remaining 90 *S*s, 21 were female and 69 were male.

TABLE 1

EXPERIMENTAL DESIGN

Adaptation	Conditioning	Extinction
All *S*s	5-sec. ISI	Informed Noninformed
	500-msec. ISI	Informed Noninformed
	Backward ISI	Informed Noninformed

Apparatus.—The recording equipment, UCS, and electrodes have been described elsewhere (Bridger & Mandel, 1965). As a further precaution against electrode polarization five sets of electrodes were used in rotation. A differential conditioning paradigm consisting of a positive (CS+) and a negative (CS−) stimulus was used; the CSs were two 25-w. bulbs mounted horizontally on a panel with centers 24 in. apart. With *S* lying supine the panel was placed 8 ft. up on a wall approximately 10 ft. from his head. Stimulus durations and ISIs were controlled by Hunter timers.

Design.—Table 1 summarizes the experimental design. Thirty *S*s were randomly assigned to each of the three ISI groups and as to whether they were to be Informed or Noninformed. The adaptation series consisted of 20 trials, 10 to each CS in a fixed random order. The conditioning series consisted of 50 trials, 25 to each CS. This 50-trial series was divided into five blocks of 10 trials with each block consisting of three CS-UCS pairs (CS+), two unreinforced CS+ test trials, and five presentations of the unreinforced light (CS−). Therefore, 15 of the 25 CS+ presentations were reinforced yielding a 60% partial-reinforcement procedure. A different random order of presentation of CS+ and CS− was used within each block of 10 trials with the restriction that no two test trials (nonreinforced CS+) occur adjacently. For the backward conditioning group the UCS was followed by a 1-sec. delay after which the CS+ appeared for 500 msec. For the two forward ISI groups a delay conditioning paradigm was used with the onset of the UCS coterminous with the offset of the CS. The reinforced trials were No. 1, 3, 10, 13, 18, 19, 23, 26, 29, 31, 35, 41, 42, 46, and 49. The nonreinforced test trials to CS+ were No. 6, 8, 12, 15, 21, 27, 32, 37, 44, and 50. The temporal duration of the UCS on the first reinforced trial was 200 msec.; on each succeeding reinforced trial the temporal duration was increased 100 msec. until the maximum of 500 msec. was reached on the fourth reinforced trial. The extinction series consisted of 20 trials, 10 to each CS randomly presented. This general design is a modification of that reported by Grings and Lockhart (1963).

The intertrial interval varied from 10 to 25 sec. Since a GSR is recorded as fall in skin resistance, a trial was not presented unless *S*'s skin resistance was either increasing steadily or remained stable for a few seconds.

Procedure.—The general description of the experimental situation and the procedure for electrode attachment have been described elsewhere (Bridger & Mandel, 1965). Each *S* was informed that the experiment would contain three parts: (*a*) a period when his responses to the lights alone, as they flashed on and off would be recorded; (*b*) a period when the light would be "associated with" electric shock; (*c*) a final period during which his responses to the lights alone would again be recorded. The adaptation series was then presented and at the end of this series *E* returned to *S*'s room and reiterated that there would be two more parts to the experiment; one part during which shocks would be given and a second part during which no shocks would be given. The CS+ light was then designated and *S* was told that shocks "would be associated with" this light but not the other (CS−). The *S*s in the Noninformed groups were told that *E* would not return to the experimental room until the experiment was completely over. The conditioning series was then presented and for the Informed *S*s, *E* returned at the end of the conditioning series, removed the UCS electrodes, wiped *S*'s leg clean of electrode jelly, and reassured him that no further shocks would be given. These Informed *S*s were told that the purpose of the final part of the experiment was to see what effect the shocks had on their responses to the lights. The extinction series was then presented. Since the purpose of this procedure was to clearly delineate acquisition from extinction for the Informed groups, an analogous interruption was *not* provided for the Noninformed groups.

Postexperimental inquiry.—Verbally communicated information of no further shocks and removal of the UCS electrodes would seem to be sufficient to remove *S*'s cognitive expectancy of further reinforcement. However, there always exists the problem of *S*'s degree of belief in the information given to him by *E*. To ensure that only the protocols of those *S*s who believed *E* would be included in the Informed groups, they were given a questionnaire containing the following question, "After you were told that no more shocks would be given and the electrodes were removed, how sure were you that you would not be shocked?" Below this statement appeared a horizontal 10-cm. line at one end of which was the phrase, "Positive no shocks"; at the other end, "Positive more shocks." The *S* was instructed to indicate by a vertical line his degree of belief or disbelief. The record of any *S* who did not mark the absolute end of the line indicating "Positive no shocks" was eliminated from the sample. In addition, those *S*s who indicated complete belief in the extinction instructions were specifically asked whether it had at all occurred to them that they might be shocked via the GSR electrodes. Only those who answered this final question negatively were accepted in the final sample. Thirty-five *S*s were discarded by this procedure, amounting to 37% of the total sample of Informed *S*s. Of these 35, 8 were from the 500-msec. ISI group, 9 from the 5-sec. group, and 18 from the Back group.

Response criteria and measurement.—A GSR was defined, for all groups, as the maximum decrease in skin resistance of any response beginning from 1 to 5 sec. after the onset of a CS. The 5-sec. cutoff point was chosen to eliminate potential offset orienting responses to the long CS in the 5-sec. ISI group. The measurement procedure has been described elsewhere (Bridger & Mandel, 1965) with the final score assigned to a response computed as log (ΔC+1), i.e., log conductance change.

Conditioning criterion.—The present design provided for five CS− presentations for each two nonreinforced CS+ test trials within each block of 10 conditioning trials. The choice of which particular CS− presentation was designated as the control stimulus for comparison purposes was therefore arbitrary. The present procedure followed that of Lockhart and Grings (1964) and designated the CS− presentation most closely preceding, but not necessarily adjacent to the test trial as the comparison stimulus.

RESULTS

Acquisition.—A series of *t* tests were performed to test for a difference be-

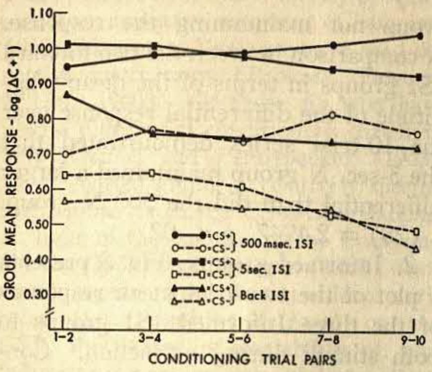


FIG. 1. Two-trial mean conditioning test-trial responses of the three ISI groups.

tween the Informed (I) and Noninformed (N) groups within each ISI in their mean response levels to CS+ and CS−. There were no significant differences, $p > .20$ ($df = 28$) all comparisons; therefore the I and N groups within each ISI were combined for further analyses. Figure 1 presents a plot of the two-trial mean response levels of the three ISI groups to both stimuli on the acquisition test trials. The *t* tests for differences between the 10-trial mean CS+ and mean CS− levels demonstrated that all ISI groups responded differentially, $p < .001$ ($df = 14$), all comparisons. An analysis of variance of the 10-trial mean differential responses (i.e., mean CS+ minus mean CS−) demonstrated significant differences between the groups in the size of the differential, $F\ (2, 87) = 6.1555$, $p < .01$. Further analysis using the Duncan multiple-range test (Duncan, 1955) indicated greater differential responding in the 5-sec. ISI than in either of the other two ($p < .01$, both comparisons). The 500-msec. ISI and Backward ISI groups did not differ significantly ($p > .10$).

An analysis of variance of the 10-trial mean response levels to CS+ indicated significant differences between the groups, $F\ (2, 87) = 3.5301$, $p < .05$. A Duncan test failed to demonstrate significant differences between the two forward ISI groups ($p > .10$); however both groups were significantly higher than the Back ISI group ($p < .05$). Separate analyses of the first and second halves (i.e., first five and last five test trials) of the acquisition series indicated that over the first five test trials there were no significant differences between the groups in terms of their mean CS+ levels, $F\ (2, 87) = 2.1928$, $p > .05$. There were, however, significant differences in the mean CS+ levels over the last five test trials, $F\ (2, 87) = 3.9607$, $p < .05$. A Duncan test indicated that the Back ISI group had a lower response level than either of the two forward ISI groups ($p < .05$, both comparisons). There was no significant difference between the two forward groups ($p > .05$).

An analysis of the 10-trial mean CS− response level indicated significant differences among the ISI groups, $F\ (2, 87) = 3.2612$, $p < .05$. The Duncan test demonstrated that the 500-msec. ISI group was significantly higher than the Back ISI group ($p < .01$), and that there was a tendency for it to be higher than the 5-sec. ISI group ($p\ .05 < R < p\ .10$). The first and second halves of the acquisition

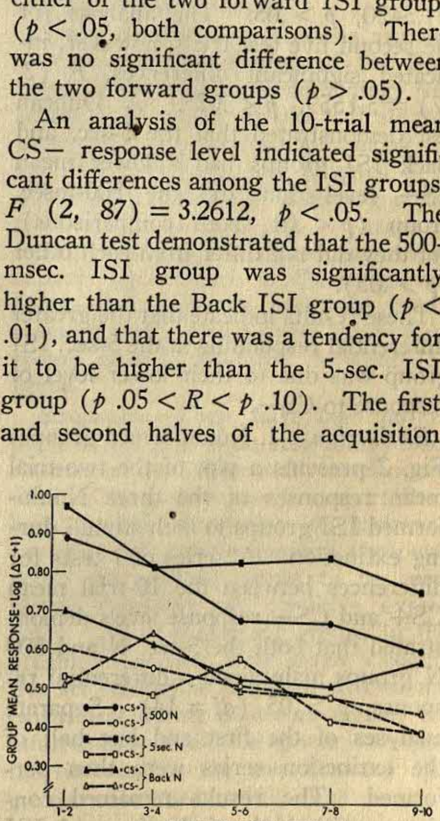


FIG. 2. Two-trial mean extinction responses of the Noninformed groups.

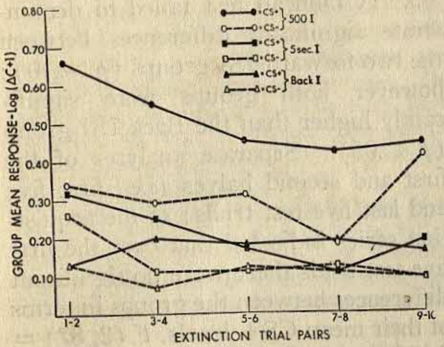


FIG. 3. Two-trial mean extinction responses of the Informed groups.

series were then analyzed separately. No significant differences were found for the first five test trials, $F\ (2, 87) = 1.5420$, $p > .05$. The analysis for the second five test trials, however, indicated significant differences, $F\ (2, 87) = 5.1559$, $p < .025$. A Duncan test demonstrated that the 5-sec. and Back ISI groups had a lower mean CS$-$ level than did the 500-msec. group ($p < .01$, both comparisons), but they did not differ from each other ($p > .05$).

These results indicate that the greater differential responding in the 5-sec. ISI group was due to their lower level of response to CS$-$.

Extinction.—1. Noninformed groups: Fig. 2 presents a plot of the two-trial mean responses of the three Noninformed ISI groups to both stimuli during extinction. A series of t tests for differences between the 10-trial mean CS+ and CS$-$ response levels demonstrated that both the 5-sec. N and 500 N groups maintained a differential response, $p > .05$ ($df = 14$). Separate analyses of the first and last half of the extinction series were then performed. The results remained consistent with both of the forward ISI groups demonstrating continued differential responding and the Back N group not maintaining the response. A comparison between the two forward ISI groups in terms of the mean magnitude of the differential response over the 10-trial series demonstrated that the 5-sec. N group maintained a larger differential than did the 500 N group, $t\ (28) = 2.4947$, $p < .02$.

2. Informed groups: Fig. 3 presents a plot of the two-trial mean responses of the three Informed ISI groups to both stimuli during extinction. Considering the two halves of the extinction series separately, a series of t tests demonstrated that: (*a*) over the first five extinction trials a significant differential response was present in both the 5-sec. I, $t(14) = 2.5759$, $p < .05$, and 500 I, $t(14) = 4.3912$, $p < .001$, groups but was not present in the Back I group ($p > .05$); (*b*) over the last five extinction trials no ISI group demonstrated a significant mean differential response.

A comparison between the two forward ISI groups in terms of their mean differential response over the first five trials demonstrated a larger differential response in the 500 I than in the 5-sec. I group, $t\ (28) = 3.2259$, $p < .01$. This larger differential response in the 500 I group was due to a significantly higher level of response to CS+, $t(28) = 2.6692$, $p < .02$. There was no significant difference between the two forward ISI groups in their mean response level to CS$-$ ($p > .20$).

DISCUSSION

The present research contained two variations of procedures used in previous reports of GSR conditioning. The first of these was the verbal communication to *S* of the exact stimulus relations. There has been a high degree of variability in the information given to *S* by different *E*s in previous research. Some *E*s have ignored this variable and failed to report on instructions (Champion & Jones, 1961, 1962); others have either merely provided a cursory summary statement (Champion, 1962; Silverman, 1960) or have used deceptive instructions (Prokasy, Fawcett, & Hall, 1960). Of those *E*s who have used a differential conditioning paradigm only Kimmel and Pennypacker (1963) have reported giving *S* relatively specific instructions as to the CS-UCS relations. In light of the early work of Cook and Harris (1937) and Mowrer (1938) demonstrating the effect of instructional variables (i.e., cognitive factors) on the GSR, the degree of variability of, and failure to report on, verbal instructions is surprising. Kimmel and Pennypacker noted that their own prior research suggested that failure to provide specific stimulus instructions produced a reduced degree of differential responding. When *S* is forced to make a cognitive assessment of the relation between stimuli under either deceptive instructions or no instructions, group conditioning curves may well represent a confounding between the varying rates at which different *S*s arrive at the proper conclusions and the course of true conditioning. The present experiment was designed to examine conditioning performance with the factor of cognitive assessment of the stimulus relations held constant.

The second procedural variation was the general instructional set given to all *S*s, i.e., the knowledge that there would be three parts to the experiment. Previous experiments and pilot work for the present experiment clearly indicated that this set was necessary to convince the Informed *S*s that no further shocks would be given after the UCS electrodes were removed. Without this general instructional set, few *S*s demonstrated any confidence in *E*'s veracity. When this factor became evident, the rule of experimental consistency dictated that the same set of instructions be given to the Noninformed *S*s as well.

Acquisition.—The present results support the findings of Kimmel and Pennypacker (1963) in GSR conditioning, and Hartman and Grant (1962) in eyelid conditioning, that longer ISIs produce greater differential responding due primarily to the decreased level of responding to the negative stimulus and a growth of inhibition over trials. Although the present experimental design closely resembles that of Lockhart and Grings (1964) the results do not support either their findings of maximal differential responding at the first point of comparison (i.e., after two differential reinforcements), or of constant discrimination performance throughout acquisition with no difference between the long and short ISI groups. There were, however, obvious procedural differences between the two experiments. In terms of the response to the positive stimulus, the present results confirm the findings of Kimmel and Pennypacker that the superiority of the 500-msec. ISI is not manifestly obvious although the data for the later trials were moving in this direction.

There have been no previous attempts at backward conditioning of the GSR using a *differential* conditioning paradigm. The present results indicate that when the cognitive relations among the stimuli are verbally communicated to *S*, differential backward conditioning of the GSR can be demonstrated. In terms of the response to CS+, the present results support the findings of Champion and Jones (1961) that a backward conditioned group attains a lower level of response than a 500-msec. forward ISI group.

Extinction.—There have been no reports of ISI effects on resistance to extinction of the differentially conditioned GSR. The present results support Pavlov's (1928) description of backward conditioning as an "evanescent phenomenon"; the response seems to be completely dependent on cognitive expectancy. All Noninformed groups in the present experiment were given a cognitive set relating to the third part of the experiment during which no further reinforcements would be given. Despite this information the two forward ISI groups maintained a differential response over the extinction series. This result was not true of the Back N group; after a few nonreinforced extinction trials the differential response was lost, indicating that these responses were highly susceptible to purely cognitive influences. This interpretation is underscored by the results of the Informed extinction procedure whereby removal of the shock electrodes produced virtually complete and immediate abolition of the responses of the Back I group. No more than 26% of the Back I group responded to either stimulus on any single extinction trial. The two Informed forward ISI groups however, continued to demonstrate a differential response despite the contrary cognitive expectancy of no further UCS presentations.

The maintenance of a differential response contrary to a cognitive expectancy is considered to be evidence for a level of GSR conditioning analogous to Mowrer's (1938) concept of "true learning." There was, however, an obvious interaction between this level of conditioning and the ISI parameter. There are two possible explanatory mechanisms underlying the interaction.

The first of these postulates that the amount of "true learning" is inversely related to the length of the ISI. In the present experiment the greater resistance to extinction by the 500I group may be considered as simply indicative of the greater degree of "true learning" present in this group. The higher CS− level of the short ISI group during acquisition therefore would be attributed to greater stimulus generalization produced by more "true learning."

The second possible explanatory mechanism would postulate that the degree of "true learning" was equal for the two forward ISI groups. However, as proposed by Hartman and Grant (1962), longer ISIs are more favorable for the manifestation of response inhibition regardless of whether the inhibition is a function of either the perceptual or reaction systems. The data of the present experiment lend themselves equally well to this interpretation. During acquisition, the lower CS− response level of the 5-sec. ISI group can be accounted for by the greater amount of time available for this group to respond to the negative aspect of the stimulus. For Informed extinction the counter expectancy of no further reinforcements is best able to inhibit the CS+ response in the long ISI group in that there is more time for it to act during each trial.

The data of the present experiment lend themselves equally well to explanation by both of these mechanisms. An experiment designed to analyze both shape and latency of GSR responses should clarify this issue.

REFERENCES

BRIDGER, W. H., & MANDEL, I. J. Abolition of the PRE by instruction in GSR conditioning. *J. exp. Psychol.*, 1965, **69**, 476–482.

CHAMPION, R. A. Stimulus-response contiguity in classical aversive conditioning. *J. exp. Psychol.*, 1962, **64**, 35–39.

CHAMPION, R. A., & JONES, J. E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–61.

CHAMPION, R. A., & JONES, J. E. Drive level (D) and extinction in classical conditioning. *J. gen. Psychol.*, 1962, **67**, 61–67.

COOK, S. W., & HARRIS, R. F. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GRINGS, W. W., & LOCKHART, R. A. Effects of "anxiety-lessening" instructions and differential set development on the extinction of the GSR. *J. exp. Psychol.*, 1963, **66**, 292–299.

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

KIMMEL, H. D., & PENNYPACKER, H. S. Differential GSR conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1963, **65**, 559–563.

LOCKHART, R. A., & GRINGS, W. W. Interstimulus interval effects in GSR discrimination conditioning. *J. exp. Psychol.*, 1964, **67**, 209–214.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

MOWRER, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

PAVLOV, I. P. *Lectures on conditioned reflexes.* (Trans. by W. H. Gantt) London: Morton Lawrence, 1928.

PROKASY, W. F., FAWCETT, J. T., & HALL, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of the interstimulus interval. *J. exp. Psychol.*, 1962, **64**, 513–518.

SILVERMAN, R. E. Eliminating a conditioned GSR by the reduction of experimental anxiety. *J. exp. Psychol.*, 1960, **59**, 122–125.

SWITZER, S. A. Anticipatory and inhibitory characteristics of delayed conditioned responses. *J. exp. Psychol.*, 1934, **17**, 603–620.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WICKENS, D. D., GEHMAN, R. S., & SULLIVAN, S. N. The effect of differential onset time on the conditioned response strength to elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

(Received May 7, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 44–49

EVALUATION OF A RESPONSE SPEED MEASURE OF INCENTIVE VALUE OF REWARD[1]

CHARLES Y. NAKAMURA AND BRUCE R. KRUDIS

University of California, Los Angeles

Response speeds can be obtained under experimental procedures in which S is alerted or in which S is not aware that speed scores are being recorded. The 2 types of scores may give different information. The present experiment was designed to provide evidence for the reliability and utility of a measure of the 2nd type as an indicator of attitude toward, or incentive value of, reinforcers. The results were consistent with predictions based on previous findings. The response-speed curve over trials for a high reward group leveled off after an initially rapid rise while that for the low reward group took the form of a shallow inverted U. Consequently, the high reward group gave the faster terminal response speeds. It was concluded that this type of measure has potential for investigating motivational variables in performance.

Response speed has been used as a dependent variable in experiments with humans under conditions in which *S* is alerted that speed of response is relevant to the task at hand (Elliott, 1965; Owen, 1959) and conditions in which *S* is not aware that speed is of interest to *E* (Birch, 1964; Kobasigawa, 1965; Ryan, 1965). Tecce and Tarnell (1965) referred to such speed measures as "focal movement time" and "incidental movement time," respectively. The former was the reaction time to press a key under standard reaction-time instructions and the latter the time taken to move the hand back to a handrest key. The results of Tecce and Tarnell in which these measures were obtained concurrently, and the comparison of experiments that used one or the other type suggest that the measures may provide different kinds of information about the effects of such motivational variables as incentive value and arousal level. If *S*s are told that reaction time is important, it seems reasonable that there would be a direct relation between speed and the motivational variables. Consistent with this, Owen (1959) obtained faster RTs under high motivation instructions than under relaxed or standard RT instructions. Within a multivariable experiment, Elliott (1965) found that overall response and incentive magnitude were directly related.

In procedures where response speed is not perceived by *S* to be pertinent to the task, *S*'s attention is typically focused on some aspect of performance on the focal task such as instructed response speed or learning proficiency while a response-speed measure that is incidentally related to that task is recorded without *S*'s knowledge. Consequently, the latter score is less subject to self-instruction by *S*. Birch (1964) has pointed out the potential utility of a speed measure obtained under conditions in which the explicit verbal control of the response was unlikely.

An incidental response-speed measure (IRS) has also been shown to be affected differently than a measure of activity level as a function of incentive

---

[1] This investigation was supported (in whole) by Public Health Research Grant No. HD-00908, from the National Institute of Child Health and Human Development.

conditions (Nakamura & Ellis, 1964). Procedures were used in which *S*'s attention was focused on the focal lever activity task and on the amount of reward received, while *S* had no knowledge that the time between the end of each trial and the press of a button that revealed the reward score was being recorded. Children in a low reward treatment group showed a continuous increase in *activity* level over trials while the *incidental response speed* in the corresponding trials rose rapidly in the early stages and then gradually decreased over the later trials. This suggested that effort might be distinguishable from attitude associated with continuous receipt of low rewards. That is, while work at the focal task continued at a high level, the gradual decrease in speed to make the response that produced the reward could be interpreted as reflecting discouragement or loss of interest. In contrast, *S*s given high rewards quickly attained a relatively fast speed but did not show the subsequent decrement. Similar results on IRS were obtained with college women *S*s (Nakamura & Boroczi, 1965).

The purpose of the present experiment was to further investigate the reliability of the preceding findings. The following predictions of response speed were based on that work.

1. For *S*s given High Reward (HR) the trend of response speed over trials should exhibit an intital rise followed by a gradual leveling.

2. The corresponding trend for Low Reward (LR) *S*s should be an initial increase followed by a gradual decrement. The initial rise would occur as a reaction to frustration. The decrement would reflect discouragement or reduction in incentive as the LR continued. A No Reward (NR) control group would likely yield a fairly flat curve.

3. As a consequence of 2 above, the HR treatment should produce faster terminal response speeds than the LR treatment.

Learning scores were obtained but, as in the previous work, it was expected that they would not be easily interpretable because there was no systematic contingency between correct problem solution and reward value.

METHOD

Subjects.—The *S*s were 42 women from an introductory psychology course. Fourteen *S*s were assigned to each of three reward treatment groups.

Apparatus.—A panel of lights was used to present the stimuli for a problem-solving task that involved comparisons of stimulus patterns. The panel was a 14-in. high, 20-in. wide plyboard painted white. It contained a row of five 1-in. diameter red reflectors that covered 7 w. bulbs. These were spaced 3 in. apart. Each light was identified by a 1½-in. high number glued beneath it. The numbers were 1 through 5 running left to right. A sixth unnumbered light was spaced 6 in. to the right of the fifth one. It served to tell *S* whether a response was correct or incorrect. The five stimulus lights could be flashed in different orders to form different sequence patterns. The different patterns were programmed on a punch tape. The panel rested on a table directly in front of *S* with the lights being about eye level.

The rewards were presented on a second display panel comprising a row of nine 1-in. diameter green reflectors with a 2-in. high number above each light. Each reward score was given by illumination of one of the nine lights. The greater the magnitude of the number above a light, the higher the reward score. The range of these numbers was different for the *relatively high* and the *relatively low* reward treatments. The numbers were such that the HR and LR groups both received the same absolute magnitude of mean reward but each received different relative amounts as a function of the context in which *S* saw the reward scores. In the HR treatment the numbers above the lights ranged from 1 through 9, while in the LR condition the numbers were from 5 through 13. Since both the HR and LR groups were given identical schedules of reward scores with a mean score of 7 and range of 5 through 9, the mean of 7 would be perceived as a relatively high score in the range of 1–9 (HR condition) and it would be seen as a relatively low score in the range of 5–13 (LR condition). This relative reward procedure was developed (Nakamura & Ellis, 1964) in consideration of its apparent advantages over the use of different absolute magnitudes for high and low reward treatments. That is, the variance associated with preexperimental individual differences in incentive values of the different absolute magnitudes of the reinforcer is presumed to be reduced by having all *S*s receive the same absolute magnitude. The effectiveness of this relative reward method was demonstrated in previous experiments (Nakamura & Boroczi, 1965; Nakamura & Ellis, 1964). The row of reward display lights was located at the level of the base, but to the right, of the stimulus panel. The *E* controlled the schedule of reward scores from an adjoining room.

On the table directly in front of *S* was a telegraph key. The onset of a yellow light mounted on the key signaled *S* that he could press the key to see his reward score. The time between light onset and *S*'s press of the key was recorded to .01 sec. on a Standard Electric clock.

Procedure.—The basic design involved three treatment groups: High Reward (HR), Low Reward (LR), and No Reward (NR). For the latter, pressing the key produced no light on the reward display panel. These *S*s were told that pressing the key recorded their score in *E*'s room. Before the experiment proper, all *S*s were given part II (analogies) of the Terman concept mastery test for a check on any differences in "brightness" between the groups.

The learning problem was presented by flashing the five numbered lights on the stimulus display panel in a predetermined order. In each trial three different orders (patterns) were presented with a 2-sec. interval between patterns. The first two were sample patterns and the third was the test pattern. The test pattern was more similar to one than the other of the sample patterns. After seeing the three patterns in succession, *S*'s task was to respond with the word "one" or "two" depending upon which of the two sample patterns *S* thought the test pattern resembled most. Following *S*'s response the signal (sixth) light on the panel was flashed either once or twice to indicate the correct answer. The order of presentation of these fixed triads of patterns was randomly varied within blocks of 10 trials with the restriction that no triad could appear more than once in one block. A total of 10 triads was used. It was possible for *S* to maximize correct choices by learning a particular principle that enabled identification of the correct match.

The *S*s in the HR and LR groups were instructed that at the end of *every fourth* trial the yellow light on the telegraph key would light and that as soon as it lit the key could be pressed in order to see the performance score for the previous four trials. The interval between light onset and key press was the main dependent variable in the experiment. The NR *S*s were flashed the correct responses but received no reward to indicate how well they had done. They were told that pressing the key recorded their scores in *E*'s room.

All *S*s were told that their scores depended most upon the number of correct answers they gave but that bonus points were added depending upon how soon they gave their answer and upon differential weightings of certain problems. Actually, the reward magnitudes were predetermined and were identical for all *S*s. The above three scoring contingencies were described to *S*s to render the scores more plausible in order to minimize suspicion of the scoring system. (After the experiment, the system was explained to *S*.) Each *S* received 100 trials and the *S*s in the HR and LR groups received 25 reward scores.

RESULTS

The Terman concept mastery scores showed that the treatment groups did not differ on intellectual brightness. Furthermore, as might be expected in a college sample, the correlation between these scores and the learning scores was not significant, $r = -.04$.

The three reward groups did not differ on total problem-solving learning scores. The means were practically identical. Neither did they differ by the end of training (last 20 trials), $F(2, 39) = 1.44$.

The statistical analyses of response speeds were applied to the reciprocals

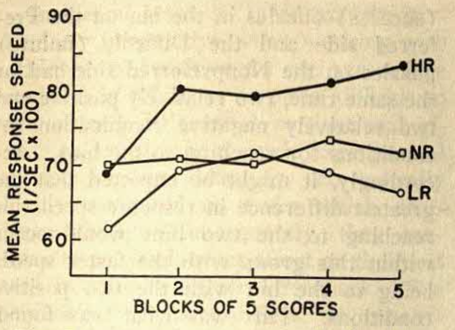


FIG. 1. Mean response speeds for three reward groups. (One response-speed score was obtained every fourth learning trial, thus giving a total of 25 scores over the 100 learning trials.)

of the time intervals between the onset of the signal light and *S*'s pressing of the telegraph key to see the reward. Overall analyses of variance tests were followed by specific comparisons. The differences between the combinations of pairs of the three treatment groups were tested for significance at the 5% level by the Duncan new multiple-range test.

The mean response speeds of the three groups plotted over Blocks of five response-speed scores are presented in Fig. 1. The Blocks × Reward analysis of variance gave significant effects for Blocks, F (4, 156) = 8.00, and for the Blocks × Reward interaction, F (8, 156) = 5.33. The Reward effect approached significance, F (2, 39) = 3.00, $p < .10$. The source of the interaction can be seen in the curves of Fig. 1. The HR group shows the expected rapid initial increase and then a gradual leveling while the LR curve also rises initially but gradually descends after the third Block. The NR group remained fairly level over the Blocks. Thus, the shapes of the curves are consistent with Prediction 1 and Prediction 2.

Trend analyses provided more complete tests of Predictions 1 and 2. Since the curve for the NR control group did turn out to be relatively flat the other curves could be compared against it. Prediction 1 would receive support if the HR curve differed from the NR curve in linear trend. In the test of the linear component the overall Mean effect was significant, F (1, 39) = 8.82, and the Reward effect approached significance, F (2, 39) = 2.73, $p < .10$. The specific difference between the HR and NR groups was significant, thus, supporting Prediction 1. The difference between the LR and NR groups was not significant.

Prediction 2 of an inverted U-shaped curve for the LR group would receive further confirmation if the LR curve differed from the NR curve in quadratic trend. The test of the quadratic component yielded a significant overall Mean effect, F (1, 39) = 9.32. Reward was not significant. However, consistent with Prediction 2, the specific comparison showed a significant difference between the trends of the LR and NR groups. The differences between the HR vs. LR and the HR vs. NR were not significant. Thus, the results of the trend tests and the curves in Fig. 1 provide support for Predictions 1 and 2.

The analysis of variance applied to the terminal speed scores (mean of Block 5) yielded a significant Reward effect, F (2, 39) = 5.00. The specific comparison test showed that the HR differed significantly from the LR group, thus confirming Prediction 3 that the HR group would, by the end of training, respond faster than the LR group. The remaining comparisons showed that the HR differed from the NR but that the LR did not differ significantly from the NR group. The Reward effect was not significant for the Block 1 scores.

DISCUSSION

The results supported the prediction that HR *S*s would exhibit an initial rise in response speed followed by a gradual leveling over trials while the LR *S*s would show the initial rise followed by a gradual decrement. The trends were quite similar to those found in previous experiments (Nakamuri & Boroczi, 1965; Nakamuri & Ellis, 1964). However, since all of these experiments used essentially the same relative reward procedure, it is possible that the reliability holds over different ages of *S*s (children and college women) but not necessarily over different reward procedures. Furthermore, the procedure involved a condition in which there was no systematic correspondence between the accuracy of performance and positive or negative reinforcement. One indirect check of the generality across reinforcement conditions is available by comparing the present results with one part of those in the experiment by Birch (1964) which used a somewhat comparable incidental response-speed measure under differently defined positive and negative conditions. Birch induced failure or success by using easy and difficult puzzles. This insured a direct relation between performance and a positively or negatively reinforcing experience. A second factor that determined that the condition was positive or negative was whether the bin which contained easy or difficult puzzles was on the preferred or nonpreferred hand side of *S*. The response speed score was the time between *S*'s (college male) moving a switch on each trial that actuated a light above one of the two bins and the moment that his hand entered the lighted bin in reaching for the puzzle that was to be solved in the subsequent trial. Combinations of Easy and Difficult tasks being on Preferred and Nonpreferred sides were counterbalanced by groups of *S*s. Consequently, response speeds for both a clearly positive and a clearly negative condition were obtained on one of the groups of *S*s. That is, as described by Birch, *S*s in the group that had the Easy (success) puzzles in the bin on the Preferred side and the Difficult (failure) puzzles on the Nonpreferred side had, at the same time, two relatively positive and two relatively negative combinations of conditions for reaching to the bins. Accordingly, it might be expected that the greatest difference in response speeds for reaching to the two bins would occur within this group with the faster speeds being to the bin with the two positive conditions. This was what was found. However, the main interest for purposes of the present comparison would be in the trends of the response speeds over trials. (Birch was not particularly concerned with this aspect of his data.) The response-speed curve for the positive condition (Birch, 1964, Fig. 1) showed a negatively accelerating curve similar to that in the HR condition of the present experiment. The curve for the negative condition showed an initial rise and then leveled off but it did not show a decrement in the later trials. However, Birch performed a second analysis on selected *S*s who remained after exclusion of certain *S*s who did not satisfy the success-failure experience requirements because they lacked proficiency in solving puzzles. The positive and the negative condition response speed curves for these selected *S*s (Birch, Fig. 2) showed differential trends paralleling those obtained in the present experiment for the HR and LR groups, respectively. Furthermore, the negative condition curve in Birch's Fig. 2 was the only one, of the four curves, that showed a decrement in the later trials.

It should be reemphasized that the response speeds to positive and negative conditions for Birch's *S*s were obtained on the same *S*s while those in the present experiment were obtained from two independent groups of *S*s, one having a positive and the other having a negative reward condition. In spite of these procedural differences, the conditions seemed sufficiently analogous to warrant the presentation of this comparison as further evidence of the generality of the results.

Incidental response speed may be particularly informative when working with achievement sensitive *S*s such as college students in whom it is often difficult to demonstrate differential motivational effects on learning in a single experimental session because these *S*s are typically highly motivated on tasks which challenge ability. Consequently, they continue to put forth great effort on the focal task in spite of negative feedback. The feedback may, nevertheless, be concurrently affecting less observable incentive and attitudinal aspects of behavior. It is possible that over a longer period of time than the typical experimental session of 1 hr., or less, effort or learning performance scores would also show reinforcement effects consistent with the subtle attitudinal-type changes. If so, incidental measures could be useful for anticipating behavior changes on the focal task. An example of the difference in performance on a focal task over short compared to a longer period of time is seen in the results of two early experiments by Hurlock. The first experiment (Hurlock, 1924) showed that groups given praise or reproof treatments both increased in performance more than a control group but the two did not differ from each other. In view of this unexpected absence of a differential effect, the second experiment (Hurlock, 1925) attempted "... to evaluate reproof, when used continuously over a given period of time, with praise used in a similar manner." Arithmetic addition tests of equal difficulty were administered on 5 successive days to school children in the fourth and sixth grades. At the beginning of each test session, starting with the second session, one group of *S*s was praised and another group was reproved for performance on the preceding day. The praised group showed a consistent increase in proficiency over days. In contrast, the reproved group did as well as the praised group during the first treatment session but declined in proficiency thereafter. Thus, Hurlock found that "Reproof, when first used, seemed to be about equal in value to praise but with continued use its effectiveness showed a decided decline." By generalization from the results of the present and the previous experiments on response speed, it would be expected that an incidental measure obtained during the initial administrations of praise and reproof would have detected motivational antecedents of the subsequent differential performance scores on the focal task.

REFERENCES

Birch, D. Incentive value of success and instrumental approach behavior. *J. exp. Psychol.*, 1964, **68**, 131–139.

Elliott, R. Reaction time and heart rate as functions of magnitude of incentive and probability of success. *J. pers. soc. Psychol.*, 1965, **2**, 604–609.

Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol.*, N.Y. 1924, No. 71.

Hurlock, E. B. An evaluation of certain incentives used in school work. *J. educ. Psychol.*, 1925, **16**, 145–159.

Kobasigawa, A. Observation of failure in another person as a determinant of amplitude and speed of a simple motor response. *J. pers. soc. Psychol.*, 1965, **1**, 626–630.

Nakamura, C. Y., & Boroczi, G. Effect of relative incentive value on persistence and response speed. *Child Develpm.*, 1965, **36**, 547–557.

Nakamura, C. Y., & Ellis, F. F. Methodological study of the effects of relative reward magnitude on performance. *Child Develpm.*, 1964, **35**, 595–610.

Owen, W. A. Effects of motivating instructions on reaction time in grade school children. *Child Develpm.*, 1959, **30**, 261–267.

Ryan, T. J. The effects of nonreinforcement and incentive value on response speed. *Child Develpm.*, 1965, **36**, 1067–1081.

Tecce, J. J., & Tarnell, M. Focal and incidental movement time as a function of shock arousal in humans. *J. Psychol.*, 1965, **59**, 155–158.

(Received May 6, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 50–53

EXTINCTION FOLLOWING PARTIAL DELAY OF REWARD WITH IMMEDIATE CONTINUOUS REWARD INTERPOLATED, AT 24-HOUR INTERTRIAL INTERVALS [1]

JANET A. DONIN,[2] C. THOMAS SURRIDGE, AND ABRAM AMSEL

University of Toronto

1-trial-a-day data are presented which indicate that partial delay of reward leads to greater resistance to extinction relative to continuous immediate reward, even when a 90-day suspension of training and a block of immediately rewarded trials are interpolated between partial-delay acquisition and extinction.

There is considerable evidence to suggest that partial delay of reward (PDR) affects later extinction in the same manner as partial reward (PR) training: 30-sec. PDR produces greater resistance to extinction than does continuous immediate reward (IR) whether highly massed or widely spaced trials are employed (Crum, Brown, & Bitterman, 1951; Logan, Beier, & Kincaid, 1956). It would therefore appear that delay trials in PDR acquisition function in a manner similar to nonrewarded trials in PR acquisition.

Several investigators have shown that the PRE is sustained through a block of continuous reward trials interpolated between PR and extinction (e.g., Jenkins, 1962; Ross, 1964; Theios, 1962). If the greater resistance to extinction produced by PDR (let us call this the PDRE) can be sustained through a block of interpolated IR trials, there would be added justification for regarding the PRE and the PDRE as companion phenomena.

The experiment we report attempts to answer the following question: Will the greater resistance to extinction typically found following PDR as compared with IR be sustained through an extended period of suspended training followed by IR reacquisition trials interpolated between acquisition and extinction?

METHOD

The *S*s were 20 male albino rats of the Wistar strain, approximately 100 days old, supplied by Woodlyn Farms, Guelph, Ontario.

A straight, 2½-in. wide plywood runway contained, in series, a 15-in. entry box, a 12-in. start box, a 50-in. alley, and a 15-in. goal box, all separated by guillotine doors. Plexiglas inner walls were lined with black inserts from start box to goal box. Four performance measures were taken by means of a microswitch on the start door and photocells in the alley.

All *S*s were on a 10-gm.-per-day deprivation schedule and were handled and tamed 28 days prior to the start of the experiment. Water was available at all times. There was no pretraining or preliminary exploration of the alley.

The *S*s were randomly assigned to one of two groups, an IR group (immediate reward on all trials) and a PDR group (30-sec. delay on half, immediate delay on the other half of the trials). The *S*s were run a total of 132 trials, one trial a day. Reward was 1 gm. in the form of two 500-mg. Noyes pellets. In the first phase of the experiment, 76 trials, Group IR received immediate reward on every trial, while Group PDR were rewarded immediately (I) on 50% of the trials and were delayed (D) in the goal box for 30 sec. prior to reward on the other 50% of the trials, according to the following

---

[1] Supported by the National Science Foundation (GB-3772) and the National Research Council of Canada (APT-72).

[2] Now at University of Illinois.

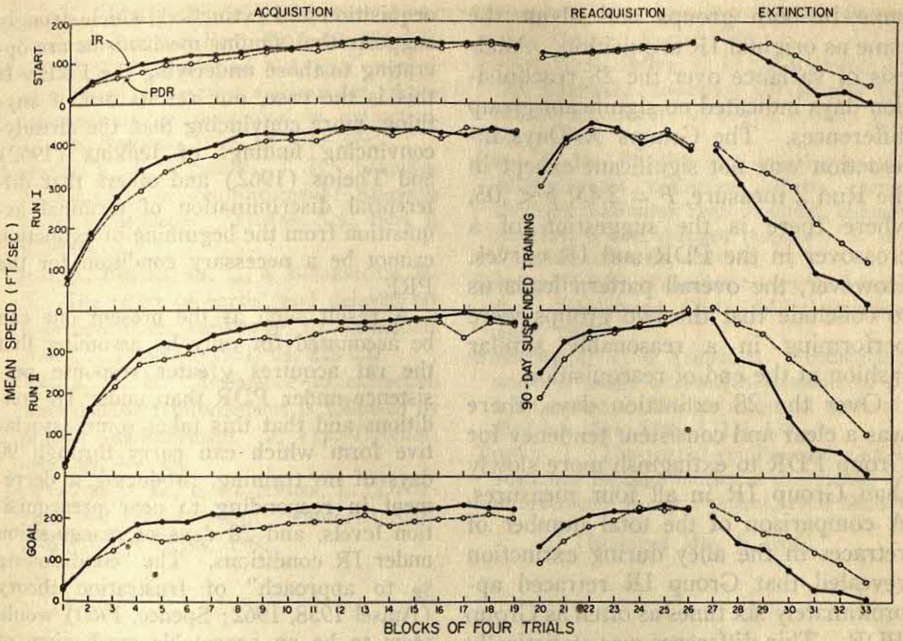


FIG. 1. Performance in the four alley segments during the Acquisition, Reacquisition, and Extinction phases.

recurring sequence: IDIDDIDIDIIDIDDI. In the second phase of the experiment there followed, in sequence, (*a*) a 90-day period during which training was suspended but all *S*s remained on the 10-gm-per-day deprivation schedule, and (*b*) a 28-day period when all *S*s got an immediate 1-gm. reward for running the alley at one trial a day. The third phase was extinction for all *S*s, still at one trial a day. If, on any extinction trial, *S* took more than 30 sec. to traverse a segment of the runway, *S* was removed from the runway and the trial was terminated. A time of 30 sec. was recorded for that measure and for each of the remaining measures of that trial. A record was kept of all trials on which *S* retraced (turned completely around) in the runway.

RESULTS AND DISCUSSION

Figure 1 presents the data of the entire experiment plotted as mean speeds (mean reciprocals of median times adjusted to ft/sec) for blocks of four trials. Included are data for four measures: Start, Run 1, Run 2, and Goal. All of the statistical analyses were carried out on the speed scores.

By the end of the original acquisition period performance of Group PDR in the Start, Run 1, and Run 2 segments is at about the same level as that of Group IR. In all of these measures there is an indication of lower average speed for the PDR group earlier in training. In the Goal measure performance of Group PDR remains inferior to that of Group IR throughout. However, analysis of variance over the 76 acquisition days revealed no significant main effect nor Group × Days interaction in any measure.

Following 90 days of suspended training both groups showed a substantial drop in performance level, the decrement being least in the start segment and greatest (relatively) in the goal segment. Reacquisition slopes were not much different from those of original acquisition, and terminal reacquisition performance was about the same in both groups and about the same as original IR acquisition. Analysis of variance over the 28 reacquisition days indicated no significant group differences. The Groups × Days interaction was not significant except in the Run 2 measure, $F = 2.43$, $p < .05$, where there is the suggestion of a crossover in the PDR and IR curves. However, the overall pattern leads us to conclude that the two groups were performing in a reasonably similar fashion at the end of reacquisition.

Over the 28 extinction days, there was a clear and consistent tendency for Group PDR to extinguish more slowly than Group IR in all four measures. A comparison of the total number of retraces in the alley during extinction revealed that Group IR retraced approximately six times as often as Group PDR. This difference was statistically significant, $t(18) = 19.68$, $p < .001$, and is taken as an indication that the PDR *S*s were more persistent in approaching the goal than the IR *S*s. Analyses of variance of the extinction speed data revealed no significant Group effect nor Group × Day interaction. However, a Group × Day linear contrast ($df = 1{,}486$) indicated that the group slopes were significantly different ($p < .01$ in the Start measure, $p < .05$ in the other three measures). As a further check on the reliability of these findings, estimates of the slopes of the running speed curves across Extinction Days 1–16 were obtained for each measure for each *S*, and were summed across measures. A comparison of these estimates indicated statistically significant differences, $t(18) = 5.16$, $p < .001$.

While the acquisition and reacquisition findings in the present study are of some interest, the main finding is that the PDRE can, in fact, be sustained through a 90-day suspended training period and a block of IR trials interpolated between acquisition and extinction, which strongly suggests that similar mechanisms are operating to those underlying the PRE. If this is the case, our results are, if anything, more convincing than the already-convincing findings of Jenkins (1962) and Theios (1962) and others that differential discrimination of terminal acquisition from the beginning of extinction cannot be a necessary condition for the PRE.

A result such as the present one can be accounted for only by assuming that the rat acquires greater response persistence under PDR than under IR conditions and that this takes some associative form which can carry through 90 days of no training, producing a decrement in responding to near preacquisition levels, and 28 days of reacquisition under IR conditions. The "conditioning s_F to approach" of frustration theory (Amsel 1958, 1962; Spence, 1960) would seem to be an acceptable mechanism of this sort, assuming that delayed reward acts frustratively like nonreward. Cognitive dissonance (Lawrence & Festinger, 1962) would also be a possible mechanism for such transfer of persistence effects. Logan's (1960) interpretation that *S*s "stop $r_G$ing" to cues following nonreinforcement is not ruled out provided it is reasonable to argue that *S*s do not make r_Gs during a 30-sec. delay period preceding reinforcement. Capaldi's (1964) account in terms of sequential effects of nonreinforcements (N-lengths) can account for the findings of the present experiment only if it can be shown that aftereffects of nonreward (*a*) carryover a 24-hr. interval in rats—the evidence here seems equivocal (Capaldi & Spivey, 1964; Surridge & Amsel, 1965)—and (*b*) that the aftereffect of nonreward on a delayed reward trial is not obliterated by the inevitable terminal reward.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning:

Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

Capaldi, E. J. Effect of N-length, number of different N-lengths, and number of reinforcements on resistance to extinction. *J. exp. Psychol.*, 1964, **68**, 230–239.

Capaldi, E. J., & Spivey, J. E. Stimulus consequences or reinforcement: Stimulus traces or memory. *Psychon. Sci.*, 1964, **1**, 403–404.

Crum, J., Brown, W. L., & Bitterman, M. E. The effect of partial and delayed reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1951, **64**, 228–237.

Jenkins, H. M. Resistance to extinction when partial reinforcement is followed by regular reinforcement. *J. exp. Psychol.*, 1962, **64**, 441–450.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement.* Stanford: Stanford Univer. Press, 1962.

Logan, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

Logan, F. A., Beier, E. M., & Kincaid, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

Ross, R. R. Positive and negative partial-reinforcement extinction effects carried through continuous reinforcement, changed motivation, and changed response. *J. exp. Psychol.*, 1964, **68**, 492–502.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N.J.: Prentice-Hall, 1960.

Surridge, C. T., & Amsel, A. Performance under a single alternation schedule of reinforcement at 24-hour intertrial intervals. *Psychon. Sci.*, 1965, **3**, 131–132.

Theios, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

(Received May 6, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 54–56

DISTRACTION AND STROOP COLOR-WORD PERFORMANCE[1]

B. KENT HOUSTON AND THOMAS M. JONES

Laboratory of Human Development, Stanford University

3 groups of 14 college students performed the Stroop Color-Word (CW) and Color-Name (CN) tasks under 3 conditions of auditory distraction: varied noises to be ignored; varied noises requiring attention for a spoken command; and no noise. It was hypothesized that there would be a specific interaction between CW and the condition of noises to be ignored. The hypothesized interaction was the only one found to be significant ($p < .05$). A case is made for labeling the significant interaction as one between inhibitory processes.

In this study, we wished to discover whether performance on a task whose performance required inhibition of a prominent response (the Stroop Color-Word test) could be improved if *S* was subjected to a variety of distracting noises which he was told to ignore. Attending to the distracting noises was conceived as a response (Solley & Murphy, 1960) which was to be inhibited. Our notion was that while the inhibitory process of ignoring the distracting noises would facilitate a task also requiring inhibition, it would not facilitate other tasks which did not require inhibition. In addition, we expected that other kinds of distraction—those involving vigilance for a cue and subsequent performance of a response to that cue—would not facilitate performance on the Stroop.

In the Stroop Color-Word (CW) test, the stimuli are the names of four colors (green, red, orange, and blue), each of which is printed in one of the other three colors. That is, the word "red" may be printed in either green, orange, or blue ink. The four color words are repeated randomly for a total of 176 color names. The *S* is asked to say the name of the color in which each word is printed. In order to be successful at this task, *S* must inhibit the response of simply reading the word.

A similar task yet which does not require inhibition of a prominent response is the Color-Naming (CN) task. This is composed of 86 sets of five colored asterisks, each set of which is printed in one of the same four colors used in CW. The *S*s were asked to name as quickly as possible the color of ink in which each set of asterisks were printed.

To assess the specificity of the effects of "inhibition distraction" on a task requiring inhibition, performance on the two experimental tasks, CW and CN, was compared across three treatment conditions: distraction allowing for blanket inhibition; distraction requiring vigilance; and quiet. The hypothesis tested was that the blanket inhibition distractor would facilitate the task involving inhibition (CW) but not the other (CN), and the vigilance distractor would facilitate neither.

PROCEDURE

Forty-five college students were randomly assigned to one of three groups: inhibition

---

[1] This investigation was supported by United States Public Health Research Fellowship MH 15,757-03 and United States Public Health Research Grant HD 01092-01. The authors wish to express their gratitude to Eleanor Maccoby for her invaluable aid in completing this study and to Karl Konrad and John Masters for their constructive commentary on the manuscript.

TABLE 1

MEAN COMPLETION TIME AND DIFFERENCE SCORES OF COLOR-WORD AND COLOR-NAMING TASKS FOR THE THREE DISTRACTION GROUPS

Tasks[a]	Groups			Difference Between Groups
	Control	Inhibition Distraction	Vigilance Distraction	
M Color-Word Time Score	183.9	172.0	182.6	C-ID* ID-VD C-VD
M Color-Naming Time Score	60.0	61.5	63.8	C-ID ID-VD C-VD
M Diff. between Color-Word and Color-Naming Time Scores	123.9	110.5	118.9	C-ID* ID-VD* C-VD

[a] In seconds.
* $p < .05$.

distraction (ID); vigilance distraction (VD); and control (C). In the two distraction groups, ID and VD, *S*s performed the experimental tasks, CW and CN, while the same, taped, distracting noises were delivered to them through earphones at a high but not uncomfortable intensity. The noises consisted of a variety of familiar sounds, e.g., trains, dripping water, etc., and unfamiliar sounds, e.g., electronic music, gibberish, etc. In ID, *S*s were instructed to completely ignore the noises. In VD, they were told to tap the table with a pencil whenever they heard the spoken word "tap" on the tape. The word "tap" was interspersed randomly but occurred on an average of about once every 11 sec. In C, *S*s performed the experimental tasks without the auditory tape.

The order in which CW and CN were performed was counterbalanced over *S*s. The time required to complete each task was the dependent measure.

RESULTS AND DISCUSSION

In order to insure the comparability of the three groups, *S*s were evaluated on grade-point average. C tended to have a lower GPA than the other two groups. Since significant correlations were obtained between the dependent measures and GPA ($r = +.37$, $p < .05$, for CW, and $r = +.412$, $p < .01$, for CN) and C contained 17 *S*s whereas the other two groups contained 14 *S*s apiece, the three *S*s with the lowest GPAs in C were dropped before the dependent measures were analyzed. With C so adjusted, there were no significant differences in mean GPA between the groups, $F\ (2, 39) < 1$.

The order of presentation of the experimental tasks did not produce a significant main effect. Interactions between order and treatment group or experimental task were also not significant.

For each group, the mean time taken to complete each task is given in Table 1. The Kruskall-Wallis *H* test (McNemar, 1962) applied to the data indicates *S*s in the ID condition took significantly less time to complete CW than did *S*s in C. Further, *S*s in ID took less time to complete CW than *S*s in VD though this result is not quite significant ($p = .07$). VD results were not significantly different from C. While there is a general increase in time to complete CN over the three groups (C < ID < VD), it is not significant. A direct test of the interaction between tasks and treatment is afforded by comparing the difference scores between CW and CN among the three groups (last line in Table 1). The mean difference score for the ID group is significantly different from those of the other two which do not differ significantly from one another.

The results thus support the hypothesis that distraction involving blanket inhibition does significantly enhance Stroop Color-Word performance yet does not do so for a similar task, CN, nor does another (vigilance) type of distraction have an enhancing effect for either task.

It should be noted that three studies have demonstrated that physiological arousal facilitates performance on the Stroop. Agnew and Agnew (1963) and Tecce and Happ (1964) created arousal by threat of electric shock and Calloway (1959) induced arousal by the administration of drugs.

Since Harmon (1933) and Helper (1957) demonstrated that physiological measures presumed to relate to arousal states are increased under conditions of distraction, we might say that in our study both the distracting conditions produced arousal, but that the facilitating effects of arousal in the VD condition were vitiated by the demand for division of attention.

In defense of the hypothesis that an inhibition interaction is operating in our study, we turn to another measure of performance on the Stroop. The usual method of scoring the Stroop is to subtract the speed of color naming in CN from that in CW. Calloway and Stone (1960) as well as Agnew and Agnew (1963) report that this difference score was *less sensitive* to arousal effects than an addition of the two scores. It appeared that arousal facilitated performance on both tasks to a small extent though CW somewhat more than CN. To subtract one score from the other vitiated the effect of arousal because one was, in effect, subtracting one increment due to arousal from another. The effect of arousal was made more salient by adding the scores and thereby compounding the increments. In Agnew and Agnew's study, the mean difference score in the arousal condition was not significantly different from that of the control condition. In our study, a significant difference between difference scores was obtained, i.e., the facilitation due to ID was specific only to the interference task. We can tentatively consider this interaction as the manifestation of another effect, one due to an interaction of inhibitory processes. Without further exploration, however, we may not state conclusively whether either or both processes were operating in this study.

REFERENCES

AGNEW, N., & AGNEW, M. Drive level effects on tasks of narrow and broad attention. *Quart. J. exp. Psychol.*, 1963, **15**, 58–62.

CALLOWAY, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. *J. ment. Sci.*, 1959, **105**, 382–392.

CALLOWAY, E., & STONE, G. Re-evaluating focus of attention. In I. L. Uhr and J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley 1960.

HARMON, F. L. The effects of noise upon certain psychological and physiological processes. *Arch. Psychol.*, 1933, No. 147.

HELPER, M. M. The effects of noise on work output and physiological activation. *U. S. Army Med. Res. Lab. Rep.*, 1957, No. 270.

MCNEMAR, Q. *Psychological statistics,* 3rd Ed. New York: Wiley 1962.

SOLLEY, C. M., & MURPHY, G. *The development of the perceptual world.* New York: Basic Books, 1960.

TECCE, J. J., & HAPP, S. J. Effects of shock-arousal on a card-sorting test of color-word interferences. *Percept. mot. Skills,* 1964, **19**, 905–906.

(Received May 12, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 57–61

EFFECT OF REINFORCEMENT SCHEDULES ON REWARD SHIFTS[1]

P. J. MIKULKA, R. LEHR, AND W. B. PAVLIK

Rutgers University

Rats given 45 runs to either 1- or 10-pellet rewards under either continuous or partial-reinforcement schedules were then given 40 runs to a 1-pellet reward under either continuous or partial-reinforcement schedules. The major findings were: (a) Standard "depression" effects occurred in both start and goal speeds following training with continuous reinforcement. (b) No "depression" effects occurred in any measure following training with partial reinforcement. (c) Shifts in schedules in the groups trained with 1 pellet had no effect on start speed measures; goal speeds declined significantly in Ss shifted from continuous to partial reinforcement and increased slightly in Ss shifted from partial to continuous reinforcement.

The present experiment evaluates the effect of partial reinforcement (PRF) on the "depression" effect. Generally, decreasing reward magnitude after extended acquisition with continuous reinforcement (CRF) produces a "depression" effect (Crespi, 1942; Gonzalez, Gleitman, & Bitterman, 1962; Vogel, Mikulka, & Spear, 1966). An extension of the Spence-Amsel Frustration hypothesis leads one to believe that this "depression" will not occur following PRF training. According to this hypothesis a classically conditioned reward expectancy (r_r-s_r chain) will evoke frustration responses when the reward on any trial is discriminably less than the expected reward (Bower, 1962). Frustration will elicit competing behaviors which produce decrements in the instrumental response below the level that would be expected from a gradual change in the underlying incentive mechanism (Pereboom, 1957). However, *S*s trained under PRF will have conditioned cues of anticipatory frustration to the ongoing instrumental response. Following a decrease in reward magnitude the frustration evoked by the lower reward produces familiar frustration cues that have already been conditioned to the ongoing response sequence. Therefore, the lowering of reward should produce no competing responses, but only a gradual decrement in response strength to the appropriate level. Thus, no depression in performance after reward reduction should be found in *S*s trained under partial reinforcement. This derivation is not unique to the frustration position, but could be derived from a number of the leading theories of partial reinforcement (e.g., dissonance theory, the discrimination hypothesis).

METHOD

Subjects and apparatus.—The *S*s were 64 naive Sprague-Dawley female albino rats approximately 50–70 days old at the start of the experiment. They were maintained at approximately 80% body weight by food deprivation throughout the experiment.

The apparatus was a straight runway with a 70-in. alley, 12-in. start box, and a 10.5-in. goal box. The start and alley units were 5 in. high × 5 in. wide. The goal box was 10.5 in wide × 5 in. high. The start box was white and the alley and goal box were gray. The entire runway had Plexiglas covers. Plexiglas doors separated the start and goal boxes from the alley. Standard

---

[1] This research was supported by the United States Public Health Service Grant MH-07844.

Electric timers and Hunter photo relays were used to time *S*'s progress through the runway. Three time measures were recorded. Start time was recorded from when the start-box door opened until *S* interrupted Beam No. 1 located 12 in. down the runway. Running time was recorded from Beam No. 1 to Beam No. 2, 6 in. from the goal-box door. Beam No. 2 activated a third, goal timer, which ran until *S* interrupted a light beam 15 in. away, 2.5 in. from the rear of the goal box and just before the food cup.

Design and procedure.—The experiment involved a 2 × 2 × 2 design, with two reward magnitudes (1 or 10 45-mg. Noyes pellets), two reinforcement schedules during acquisition (PRF-50% or CRF), and two schedules of reinforcement during the shift phase (PRF or CRF). All *S*s were run to a 1-pellet reward in the shift phase. The resulting eight groups are designated as follows: $C_{10}C$ (continuous reinforcement and a magnitude of 10 pellets in acquisition, and under continuous reinforcement during the shift phase), $C_{10}P$ (similar to the above, except that the *S*s were under a partial schedule during shift), $P_{10}P$, $P_{10}C$, C_1C, C_1P, P_1P, and P_1C.

The *S*s were fed 8 gm. per day, consisting of the food received in the test situation and a supplementary quantity of Purina chow. They were handled 3 min. a day on Days 1–3. On Days 4–5 *S*s were individually habituated to the alley and given several Noyes pellets in their home cages to familiarize them with the reinforcement to be used. On Day 6 acquisition began, and consisted of five trials per day with an average intertrial interval of 40 min. The first five trials were all reinforced for all *S*s. The partial reinforcement schedule was constructed with the restriction that three and two reinforcements be given on alternate days. A total of 9 days of acquisition (45 trials) and 8 days following the incentive shift (40 trials) were given. During the acquisition and shift phases *S*s were removed immediately after the reward was consumed; on nonrewarded trials they were confined 20 sec. in the goal box.

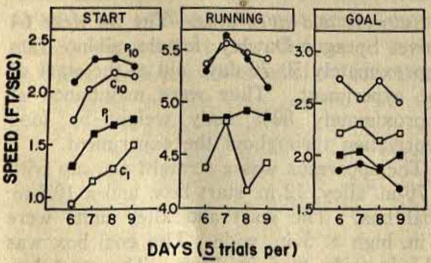


FIG. 1. Response speeds of the experimental groups over the three speed measures of the response chain on the last four days of acquisition.

All latencies were converted to speed scores (1/latency) for analysis. Analyses were performed on individual mean speed scores. The shift data (Fig. 2) are plotted in terms of the mean difference in speeds between the large reward groups and their appropriate controls. That is, the mean speed for the large reward minus the mean speed for the controls.

RESULTS

Analyses of variance of speeds over the last 4 days of acquisition yielded significant effects of reward magnitude in start and running speeds, *F* (1, 56) = 49.07 and 22.00, $p < .001$, respectively, with the large reward groups running faster than their respective small reward groups (See Fig. 1). The effect of schedules was significant in start speed, *F* (1, 56) = 7.55, $p < .01$, with the PRF *S*s starting faster than their respective CRF *S*s. This difference disappeared in the running measure, and then reversed at the goal, the CRF *S*s running faster than the PRF *S*s, *F* (1, 56) = 32.64, $p < .001$. In goal speed the CRF *S*s showed the standard magnitude effect with the C_{10} group running faster than the C_1 group. However, the PRF *S*s did not differ significantly as a function of magnitude, although the 10-pellet *S*s tended to run more slowly. This is indicated by the Magnitude × Schedules interaction, *F* (1, 56) = 9.58, $p < .01$. Start and running speed increased over the last 4 days, *F* (3, 168) = 31.16 and 41.00, $p < .001$, respectively, while goal speeds did not, *F* (3, 168) = 2.25, $p > .05$. The effect of the "dummy" variable, the to-be-introduced schedules in the shift phase, yielded *F*s of less

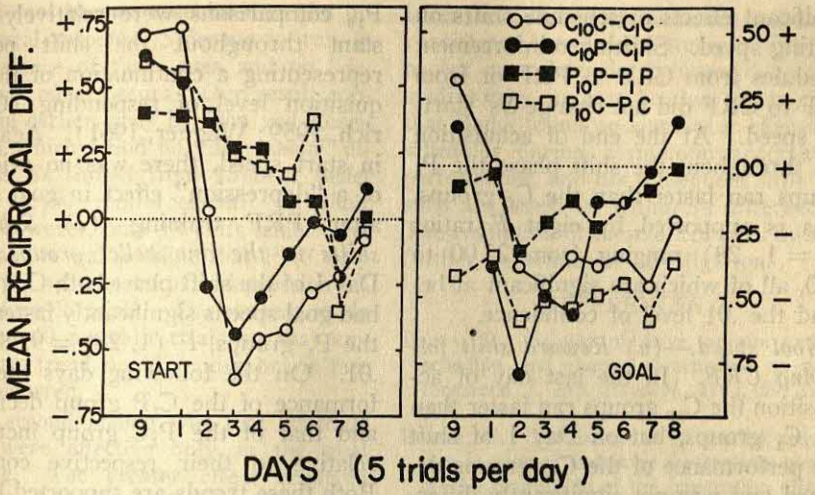


FIG. 2. Represents the mean difference in speeds between each large reward group and its appropriate control group over the last day of acquisition and the 8 days of shift. (The lefthand graph represents the start-speed measure and the righthand graph represents the goal-speed measure.)

than unity in all three response measures.

Following the Shift in Reinforcement

Three factorial analyses, one for each major comparison, were performed on each day's data following the incentive shift to track closely the performance changes over days. The three major comparisons were (*a*) the effect of reward shift on the CRF groups ($C_{10}C$, $C_{10}P$, C_1C, and C_1P), (*b*) the effect of reward shift on the PRF groups ($P_{10}P$, $P_{10}C$, P_1P, and P_1C), and (*c*) the effect of shifts in the percentage of reinforced trials on performance (C_1C, C_1P, P_1P, and P_1C). Start and goal speeds are presented separately; the results for running speeds were similar to those for start speeds and are not presented.

Start speed.—(*a*) *Reward shift following CRF.* On Day 1 of the shift phase the C_{10} groups ran faster than their respective controls, F $(1, 28) = 10.58$, $p < .01$, as they had over the last 4 days of acquisition. This is shown in Fig. 2 by the initial positive scores. Their performance (C_{10}) rapidly decreased and was significantly below the C_1 groups on Day 3, F $(1, 28) = 6.58$, $p < .05$, and approached significance on Day 4, F $(1, 28) = 3.18$. Thereafter, the C_{10} groups' performances increased slowly to the level of the controls, as indicated by the gradual approach of the negative scores toward zero (See Fig. 2). Thus, a standard "depression" effect occurred following training with CRF. (*b*) *Reward shift following PRF.* Analyses of the speed data indicated that the P_{10} groups initially ran faster than the controls on Days 1 and 2, F $(1, 28) = 11.87$ and 9.54, $p < .01$, respectively, but then gradually decreased to the level of the controls. This is seen in Fig. 2 where the positive difference scores approach zero over trials. The slight negative dip on Day 7 did not approach significance. Thus, there was *no* evidence for a "depression" effect following training under partial reinforcement. (*c*) *Schedule shifts in the one pellet groups.* There were no significant effects of schedule shifts on starting speed. Shifting reinforcement schedules from CRF to PRF or from PRF to CRF did not change *S*s' starting speed. At the end of acquisition and throughout the shift phase the P_1 groups ran faster than the C_1 groups. This is supported by eight F ratios ($df = 1, 28$) ranging from 23.00 to 9.90, all of which are significant at beyond the .01 level of confidence.

Goal speed.—(*a*) *Reward shift following CRF.* In the last day of acquisition the C_{10} groups ran faster than the C_1 groups, but on Day 1 of shift the performance of the C_{10} groups declined and was not significantly different from the C_1 groups. On Days 2 through 5 the C_{10} groups ran significantly slower than their controls, F (1, 28) = 23.18, 6.83, 6.09, and 5.42, respectively, $p < .05$. By Day 7 the $C_{10}P$ group had adjusted to the level of its control, while the $C_{10}C$ group was still running below its control group (See Fig. 2). Thus, as in start speed, a significant "depression" effect was found in the CRF groups. The effect of schedule during shift phase was highly significant throughout the 8 days of shift with F ratios (1, 28) ranging from 25.00 to 4.50, $p < .05$, with the groups on PRF ($C_{10}P$ and C_1P) showing a lower level of performance than their respective controls on CRF ($C_{10}C$ and C_1C). (*b*) *Reward shift following PRF.* When considering the effect of reward shift on the P_{10} groups, it is important to stress that the goal speed of the P_{10} *S*s was slower than that of their controls (P_1) on the last day of acquisition, F (1, 28) = 5.88, $p < .05$. The P_{10} groups were not significantly below their control groups on Day 1 of shift, but were below on Days 2–7 with F ratios (1, 28) ranging from 15.66 to 5.66, $p < .05$. Figure 2 indicates that the negative scores of the P_{10} comparisons were relatively constant throughout the shift period, representing a continuation of the acquisition level of responding (Goodrich, 1959; Wagner, 1961). Again, as in start speed, there was no evidence of a "depression" effect in goal speed after PRF training. (*c*) *Schedule shifts in the one pellet groups.* On Day 1 of the shift phase both C_1 groups had goal speeds significantly faster than the P_1 groups, F (1, 28) = 9.88, $p < .01$. On the following days the performance of the C_1P group decreased and that of the P_1C group increased relative to their respective controls. Both these trends are supported by the significant Schedules × Shift Schedules interactions over the last 6 days of the shift period, F ratios (1, 28) ranging from 29.75 to 8.42, $p < .01$. The performance level of the C_1P group equaled that of the P_1P over the last 5 days of the shift phase. Although the performance level of the P_1C group increased over the shift period, it fell considerably below the level of the C_1C group and above that of the P_1P group over the last 7 days of the shift phase.

DISCUSSION

Our major result is that the "depression" effect that follows reductions in reward with continuous reinforcement was eliminated by initial training under partial reinforcement. Implied is a similarity between nonreward and reward reduction as suggested by Bower (1962). This suggests that a similar mechanism operates in both situations, since the conditioning of cues from nonreward trials to the instrumental running response appears to have sustained performance without disruption after the reward was reduced. The proposed common mechanism is frustration (r_f—s_f) which is presumed to be aversive and which, in anticipatory form, evokes competing behaviors that cause the subsequent "depression" effect. However, if, through conditioning, frustration cues become associated to the instrumental response, *S* learns to run in the presence of these cues and no "depression" effect occurs. Our result confirms an earlier study by Ison and Rosen (1965) which found no "depression" effect for *S*s shifted from 15 to 1 pellet under partial reinforcement. In their study, however, only 18 shift trials were given and it was possible that PRF only delays the appearance of a depression effect before 18 trials. The present study provided 40 postshift trials so that any long-term trend would appear, but in fact none occurred.

Our second finding was that schedule shifts were effective only in the goal measure. The greater effect occurred with the shift from CRF to PRF, producing a large reduction in the performance to the level of the PRF control group. However, the opposite shift from PRF to CRF produced only a slight elevation in performance at the goal. This finding is consistent with the assumption that nonreward is aversive and its introduction into a reinforcement schedule will disrupt performance of a response. It is also consistent to expect this disruption to be the greatest at a point nearest the goal. These results contrast with earlier findings of Badia (1965) who found schedule shifts effective in the start and running measures but not at the goal. Badia reported an increase in start and running speed for the CRF *S*s when shifted to PRF, while our results suggested a decrease in goal speed following such a shift. He also found that a schedule shift from PRF to CRF had no effect on any of his performance measures, while in our study a slight increase in PRF *S*s' performance at the goal was found following the schedule shift. This latter fact represents the only apparent consistency between the studies. The apparent inconsistencies between these two studies may in part be due to the fact that Badia found no significant effects of schedules during asymptotic acquisition performance, and his subsequent shift results are based on within-*S*s differences. The resolution of these differences awaits further experimental investigation.

REFERENCES

Badia, P. Effects of drive, reinforcement schedule, and change of schedule on performance. *J. exp. Psychol.*, 1965, **69**, 292–297.

Bower, G. The influence of graded reduction in reward and prior frustrating events on the magnitude of the frustration effect. *J. comp. physiol. Psychol.*, 1962, **55**, 582–587.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Gonzalez, E. C., Gleitman, H., & Bitterman, M. E. Some observations of the depression effect. *J. comp. physiol. Psychol.*, 1962, **55**, 578–581.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

Ison, J., & Rosen, A. J. On changes in incentive: Some new findings and theoretical speculations. Paper read at the Psychonomic Society Meetings, Chicago, 1965.

Pereboom, A. C. A note on the Crespi effect. *Psychol. Rev.*, 1957, **64**, 263–264.

Vogel, J., Mikulka, P. J., & Spear, N. E. Effect of interpolated extinction and level of training on the "depression effect." *J. exp. Psychol.*, 1966, **72**, 51–60.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

(Received May 13, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 62–66

SERIAL LEARNING AND ORDER INFORMATION[1]

NORMAN J. SLAMECKA

University of Vermont

Serial learning was studied with varied sources of item-order information. Group 1 had temporal input order identical to prescribed output order. Group 2 had temporal input unrelated to prescribed output order, by presentation of items in spatial array independent of their temporal succession. Spatial location was the only source of order information. Group 3 also had temporal input unrelated to prescribed output, and also lacked consistent absolute position information. Relative spatial location was the only source of order information. Results showed that Group 3 was slower than Groups 1 and 2, which did not differ. Serial-position curves were sensitive to the required output order, and not to the input order. The data strongly supported a position-learning interpretation of serial learning.

The serial-learning task requires a prescribed ordering of responses in *S*'s performance (output). Some aspect of the list presentation (input), must be invariant across trials in order to communicate this desired ordering. If input were random in all respects serial learning could not take place. In the usual serial-learning task, prescribed output order is indicated by the temporal succession of items as they arrive, whether on a memory drum, or aurally from a tape recorder. Indeed, the commonly used anticipation procedure demands that output order correspond exactly to temporal input order, item by item. This reliance upon temporal succession as the sole vehicle for item-order information is not a methodological necessity. It is possible to convey order information in a controlled manner by purely spatial means, independently of temporal item succession. Sources of order information can be experimentally manipulated, and by their effects upon the task, deductions can be drawn about the processes underlying serial learning.

Research into the basis of serial learning has stressed two hypotheses (Underwood, 1963). The first asserts that associations formed between adjacent items eventuate in a chain linking all items together in their order of presentation. In this view, temporal contiguity is the key that determines which items will be linked to which others. Ebbinghaus stated, "Ideas which have been developed simultaneously or in immediate succession in the same mind mutually reproduce each other, and do this with greater ease in the direction of the original succession . . . [1964, p. 90]." The predictive power of the chaining hypothesis therefore rests upon an assumption that the associative ordering of items is determined by their order of succession in original experience.

The second hypothesis asserts that serial tasks involve some form of position learning, such that the specific locations of items are acquired. Woodworth observed, "Even when the items are exposed successively, a 'visually minded' *O* may assign each item to its place in an imaginary row or column. . .

---

[1] This research was supported by NSF Grant GB-2590. The author is grateful for the assistance rendered by Edwin Coppage. Portions of this paper were read at the Eastern Psychological Association meetings in 1966.

[1938, p. 32]." In this view the critical functional connection is between an item and its location, and not with the succeeding item. It should be recognized that there are at least two avenues to position identification. One is that an item is associated with a fixed *absolute* location, and another is that its *relative* location is established with respect to some internal reference points, or anchors.

An important theoretical distinction between the two views is that the chaining hypothesis implies that order in serial recall is critically dependent upon temporal order of arrival. Development of associative bonds through contiguity must necessarily constrain output order, since an item would be most strongly connected to the one that followed it in time. On the other hand, a position hypothesis is free from such an implication since identification of item locations, as such, does not automatically determine the order of serial recall.

Present evidence indicates that neither hypothesis (nor combinations thereof) has received consistent support in all instances (Ebenholtz, 1963a; Jensen & Rohwer, 1965; Postman & Stark, in press; Underwood, 1963; Young, 1962). The approach described below is a departure from the use of transfer designs which have characterized research on the issue thus far.[2] The information sought is whether serial learning is at all possible when temporal input order is unrelated to prescribed output order. The chaining hypothesis, based upon temporal contiguity, would say no. Further, if such learning is indeed possible, how does it compare with a condition where input and output orders are temporally identical? The serial anticipation method is theoretically ambiguous for these purposes, since it permits both chaining and position learning to eventuate in the same ordering of output. Therefore, a method was used which could convey order information by other than temporal means.

METHOD

Experiment I

Fifty-four University of Vermont psychology students served as *S*s to fulfill course requirements. Three groups of 18 each memorized a serial list of 10 CVC trigrams (2.00–2.10 m′ value from Noble [1961]), by alternate training and testing trials to a criterion of one errorless performance. Items appeared on a Lafayette memory drum at a 3-sec. rate. A training trial consisted of one presentation of the list, with *S* pronouncing each item as it appeared. A test trial consisted of a written recall in strict serial order, *S*-paced, starting with the first item and adding the succeeding ones in proper order. No skipping or later additions or deletions were permitted, and since *S* had to account for all 10 items, he put a dash wherever one could not be recalled. All *S*s were told that the list contained 10 nonsense syllables.

Group 1 had the usual presentation procedure where temporal input order corresponded to the required output order, and all items appeared in the same location in the window.

Group 2 had temporal input order unrelated to the prescribed output order. Spatial location was the only source of order information. Each item appeared in a separate fixed location in the long window of the drum. Each exposure revealed 10 dashes in horizontal array across the window, indicating positions, with the appropriate item above one of them. If the temporal succession of items is designated as 1–10, then the left-right spatial sequence was 6, 9, 5, 8, 2, 10, 4, 1, 7, 3. Thus the first item temporally was in the eighth spatial position, the second temporally was in the fifth position, and so on. With this arrangement no part of the temporal succession ever corresponded to any part of the spatial sequence. The input situation was explained to *S*, and he was instructed to order his recall according to the ***spatial*** sequence.

---

[2] Some doubts about the usefulness of such transfer designs are noted by Postman and Stark (in press), as, "... it will be difficult to arrive at a definitive identification of the functional stimulus in serial learning on the basis of transfer effects in paired-associate learning."

Group 3 also had temporal input unrelated to the prescribed output order and was, in addition, denied the use of consistent absolute position information. *Relative* spatial location was the only source of order information. The absolute locations of all items were shifted from trial to trial by displacing the entire list by a few positions. However, the spatial sequence per se, and therefore the relative positions, remained constant. In order to provide a consistent starting point for recall performance, a red frame was drawn around one of the dashes and was visible at each exposure. The frame always surrounded the same item, so that as the absolute location of that item changed across trials, so did the location of the frame. If the first-trial spatial locations in the window (numbered by temporal succession) were 3, $\boxed{6}$, 9, 5, 8, 2, 10, 4, 1, 7, then the locations on the next trial were 8, 2, 10, 4, 1, 7, 3, $\boxed{6}$, 9, 5. A total of three such shifts occurred, with the frame occupying Locations 2, 4, and 8. Recall was to be ordered by starting always with the framed item and proceeding to the right in ***spatial*** sequence. Thus for the above list, the prescribed output was $\boxed{6}$, 9, 5, 8, 2, 10, 4, 1, 7, 3.

Two arrangements of items were employed across groups, with half of the *S*s given each one. Specifically, the temporal succession of one list was the spatial sequence of the other list. All groups were thus equated with respect to the actual sequences to be learned.

Experiment II

Since Exp. I used CVCs, response integration was a factor in the task, as indicated by numerous instances of misspelled items on the early trials. It was judged desirable to determine whether the influence of varied order information would be any different when the response integration load was minimal. Accordingly, Exp. II used a list of 10 consonants. The very same conditions and procedural details were used as in Exp. I, with two exceptions. No overt pronunciation was required on training trials, and the acquisition criterion was raised to two consecutive errorless performances. There were 54 naive undergraduate *S*s, in three groups of 18 each. All *S*s were informed that the list contained 10 different consonants.

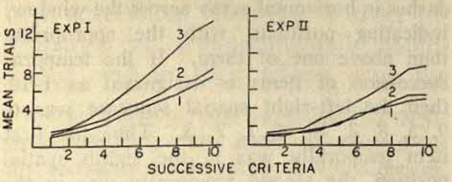


FIG. 1. **Mean trials to achieve successive performance criteria for all groups in both experiments.**

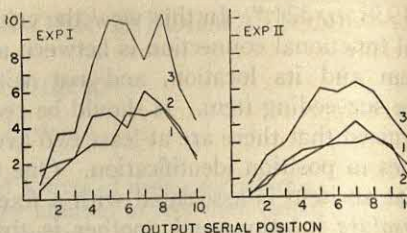


FIG. 2. **Serial-position curves for all groups in both experiments.** (Serial positions refer to the prescribed output orders.)

RESULTS

Mean total trials to acquisition in Exp. I (with *SD*s), were: Group 1—7.67 (2.39), Group 2—8.17 (2.92), and Group 3—13.50 (4.81). Duncan's range test (McGuigan, 1960) applied at the .01 level, showed that Group 3 took significantly more trials than Groups 1 and 2, which did not differ from each other, *SE* (51) = 3.64.

For Exp. II, mean trials were: Group 1—7.39 (4.05), Group 2—6.89 (1.76), and Group 3—11.06 (5.12). Duncan's test at the .01 level again showed that Group 3 took significantly more trials than the other groups, which did not differ, *SE* (51) = 4.02. In both experiments Group 3 took about 50% more trials than either of the other groups.

The course of learning under all conditions in both experiments is seen in Fig. 1, in terms of mean trials taken to achieve successive performance criteria. It is seen that at all criterion levels Groups 1 and 2 gave highly similar performances, but that the curves for Group 3 sloped upward more steeply as learning proceeded, diverging increasingly as the terminal levels were approached. Analysis of variance (Type I, Lindquist, 1953) on Exp. I indicated significance for Groups, F (2, 51) = 14.58, $p < .001$, Criteria, F (9, 459) = 194.64, $p < .001$, and the Groups × Criteria interaction, F (18, 459) = 9.98, $p < .001$. Also in Exp. II Groups, $F = 7.95$, $p < .005$, Criteria, $F = 122.19$, $p < .001$, and the Groups × Criteria interaction $F = 4.23$, $p < .001$, all had highly significant effects. The interaction found in both experiments confirms that Group 3 faced an increasingly difficult task as it approached mastery of the list. Some information as to the locus of this difficulty is gained from the serial-position curves presented in Fig. 2. The serial positions are arranged in accord with the output orders. It is seen that in both experiments, Group 3 committed relatively more errors in the middle of the list and just beyond, than did Groups 1 and 2. It is therefore reasonable to attribute the slower learning of Group 3 to the greater difficulty it encountered with midlist items. The steady divergence of Group 3 from the other groups as learning progressed (Fig. 1), reflects the fact that these particularly difficult items tended to be the last learned. The mean list position of the last learned item for Groups 1, 2, and 3, respectively, was 6.37, 6.18, and 6.32 in Exp. I, and 6.80, 6.63, and 5.82 in Exp. II.

DISCUSSION

The first question, whether serial learning is at all possible when temporal input is unrelated to prescribed output order, was clearly answered in the affirmative. Groups 2 and 3 both mastered the task in spite of the absence of temporal contiguity between adjacent items. This conclusion applies to both experiments, and is thus independent of the response integration load. The second question, concerning the comparability of such learning with the usual case where temporal input-output correspondence prevails, has two answers. When only absolute position information was available, learning was just as rapid as when temporal contiguity was additionally present, as seen in the comparisons of Groups 1 and 2. However, when the only source of order information was relative item location, learning was considerably slower, as with Group 3. The reason for this appeared to lie in the greatly increased difficulty of the middle items. If, as seems likely, the red frame was used as an anchor point (Glanzer & Dolinsky, 1965), then the middle items would have been most difficult to locate because of their greater distance, in either direction, from this anchor. Provision of absolute position information, as with Groups 1 and 2, appeared to facilitate the learning of midlist items.

The familiar shapes of the serial-position functions of Groups 2 and 3 harbor an interesting implication. The obtained primacy and recency effects corresponded to the prescribed *output* orders, and not to the input orders, which were quite different. Were order of arrival the determining factor, the curves would have different shapes entirely. Such an observation is possible only when input and output orders are experimentally separated as they were here, and it suggests that the characteristic form of serial-position curves is given to them solely by the output requirements of the task.

Since it has been shown that availability of position information *alone* can mediate serial learning, this suggests that the fundamental process in a serial task is the identification of item locations, rather than the establishment of associative links between items. It could not be plausibly maintained, for instance, that Group 2 proceeded basically by forming associations between adjacent items, since in order to have done so the locations of the appropriate items would have had to be identified *first,* and therefore position learning had to be the fundamental or prior process. Since Group 2 learned just as rapidly as Group 1, even without temporal contiguity between items, parsimony requires the conclusion that a position-learning process underlay both performances. It should also be noted that some studies have varied the absolute temporal positions of items across trials, within an anticipation task, in hopes of inducing learning by associative chaining (for instance, Ebenholtz, 1963b; Keppel, 1964). The learning was always slower when compared to the condition of constant positions across trials. However, since the *relative* item positions were still unchanged, such a method could not rule out relative position learning. The outcomes were comparable to the Group 1–Group 3 performances above, and yield to the interpretation that position learning was occurring in both conditions.

Asch, Hay, and Diamond (1960) presented serial lists in various spatial arrays, but since they used the method of serial anticipation there was no opportunity to test the effectiveness of spatial order information independently of temporal contiguity. Ebenholtz (1963a) also used a spatial array of items, but the task was paired-associates learning. The present method appears better suited for providing information about the role of spatio-temporal factors in serial learning. Since written recall tests were employed, *S* had a spatial array of items before him on each trial. Whether this may have facilitated the performances of Groups 2 and 3 is not known, but the use of spoken recall tests should give the answer.

Finally, these findings are at variance with a memory model such as Conrad's (1965), which assumes that items are stored in the order in which they are perceived, and can only be retrieved in that same order, from whatever starting point. That model is clearly inadequate for a repeated-trials situation, at any rate. It would seem that *S* is quite flexible in his ability to make use of order information, and that order of arrival is not a governing factor in serial-learning tasks.

REFERENCES

ASCH, S. E., HAY, J., & DIAMOND, R. M. Perceptual organization in serial rote-learning. *Amer. J. Psychol.,* 1960, **73,** 177–198.

CONRAD, R. Order error in immediate recall of sequences. *J. verbal Learn. verbal Behav.,* 1965, **4,** 161–169.

EBBINGHAUS, H. *Memory: A contribution to experimental psychology.* New York: Dover, 1964.

EBENHOLTZ, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *J. exp. Psychol.,* 1963, **65,** 603–608. (a)

EBENHOLTZ, S. M. Serial learning: Position learning and sequential associations. *J. exp. Psychol.,* 1963, **66,** 353–362. (b)

GLANZER, M., & DOLINSKY, R. The anchor for the serial position curve. *J. verbal Learn. verbal Behav.,* 1965, **4,** 267–273.

JENSEN, A. R., & ROHWER, W. D., JR. What is learned in serial learning? *J. verbal Learn. verbal Behav.,* 1965, **4,** 62–72.

KEPPEL, G. Retroactive inhibition of serial lists as a function of the presence or absence of positional cues. *J. verbal Learn. verbal Behav.,* 1964, **3,** 511–517.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton-Mifflin, 1953.

MCGUIGAN, F. J. *Experimental psychology: A methodological approach.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

NOBLE, C. E. Measurements of association value (a), rated associations (a′), and scaled meaningfulness (m′) for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.,* 1961, **8,** 487–521.

POSTMAN, L., & STARK, K. Studies of learning to learn: IV. Transfer from serial to paired-associate learning. *J. verbal Learn. verbal Behav.,* in press.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 33–48.

WOODWORTH, R. S. *Experimental psychology.* New York: Henry Holt, 1938.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.,* 1962, **67,** 307–313.

(Received May 13, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 67–74

CLASSICAL GSR CONDITIONING WITH CONCOMITANT EMG MEASUREMENT[1]

H. D. KIMMEL AND WILLIAM DAVIDOV
Ohio University *University of Florida*

32 college students received either paired light-shock (ISI = 5 sec.) or unpaired light and shock with GSR and EMG simultaneously recorded. During the 12 acquisition trials a significant difference between groups in GSR magnitude indicated that GSR conditioning occurred. No difference in frequency of EMG changes between groups was found, not even in 2 subgroups whose EMG was picked up from the arm whose hand received the shock. In addition, inspection of records from selected trials showed that the GSR and EMG were substantially independent in their occurrence from trial to trial. It was concluded that the proposition that GSR conditioning is an artifact of conditioning of muscular movements was not supported by these results. It was suggested that the failure to obtain EMG conditioning at all in a situation in which GSR conditioning occurred might mean that either the 5-sec. ISI was too long for EMG conditioning or that EMG conditioning can only be demonstrated by using an instructed UCR.

According to Smith (1954, 1964), classical conditioning of an autonomically mediated response, such as the GSR, occurs as an artifact of instrumental skeletal conditioning rather than as an instance of "genuine" classical conditioning of the autonomic response itself. His thesis is that, when an attempt is made to condition an autonomic response, the procedure usually involves the pairing of an indifferent stimulus (CS) with a noxious stimulus (UCS), e.g., electric shock. The *S* comes to regard the CS as a signal for the impending shock and muscularly "braces" himself against it. Supposedly, this muscular bracing is the result of reinforcement (instrumental) learning. The skeletal response is assumed to produce observable autonomic changes which are erroneously taken as evidence of classical conditioning.

The present study was intended to provide evidence relevant to Smith's proposition in a study of GSR conditioning in human *S*s. Since deep curarization of humans is infeasible and, furthermore, since Smith (1964) has rejected previous arguments based upon studies of curarized subhuman *S*s (Black & Lang, 1962), the present study involved no pharmacological intervention. Instead, a procedure known to result in "conditioning" of the GSR was used and concomitant measures of forearm muscle action potential (EMG) were obtained. It was expected that, when the EMG was picked up from the left arm with the shock electrodes on the left hand, muscular bracing in anticipation of the shock would be found. When the EMG was picked up contralaterally from the shocked left hand, however, it was assumed that there would be little evidence of anticipatory muscular activity. Both EMG-placement groups were, of course, expected to show GSR conditioning.

METHOD

Thirty-two introductory psychology students were assigned randomly to four condi-

---

[1] Under United States Public Health Service Grant M-06060-2.

tions: light-shock paired, EMG and shock ipsilateral (E-S); light-shock unpaired, EMG and shock ipsilateral (C-S); light-shock paired, EMG and shock contralateral (C-D). The *N*s for these groups were 8, 9, 6, and 9, respectively.

The GSR was picked up from the palm and back of *S*'s right hand by ¾-in. zinc electrodes in lucite cups filled with NaCl jelly. The EMG was measured from the highest point on the dorsal surface of the forearm, ⅔ of the distance from the wrist to the elbow, with the two silver electrodes 1 in. apart. The skin under the EMG electrodes was covered with Grass EEG paste. Both responses were amplified and recorded by a Grass Model 5 Polygraph (5P1-GSR and 5P3-EMG), with a paper speed of 1 mm/sec.

The equipment for delivering the CS, a white circular light, and the UCS, a 3.5-mamp. shock to the fingertips of the left hand, has been described previously (Morrow, 1967). CS duration was 5 sec., and shock duration was .1 sec., the latter beginning at the termination of the former.

All *S*s received preliminary presentations of the CS until a criterion of 2 successive nonresponses or 10 trials occurred. A minimum of 4 CS-only trials was given to all *S*s. Groups E-S and E-D then received 12 paired CS-UCS trials while Groups C-S and C-D received 12 unpaired presentations each of light and shock. Intertrial intervals for the E groups were varied unsystematically between 30 and 90 sec., while single stimulus trials for the C groups were separated unsystematically by 15–45 sec. Data were collected in a totally dark IAC (1200 series) audiometric chamber. Instructions were read via an intercom. They told *S* to relax muscularly but to remain attentive to the stimuli. Prior to removing *S* from the apparatus at the end of the conditioning session, *E* instructed him to tense the arm containing the EMG electrodes, to move his fingers, and to lift the arm. At this time it was observed that the EMG electrodes were appropriately located and, in fact, were sensitive to all such activity.

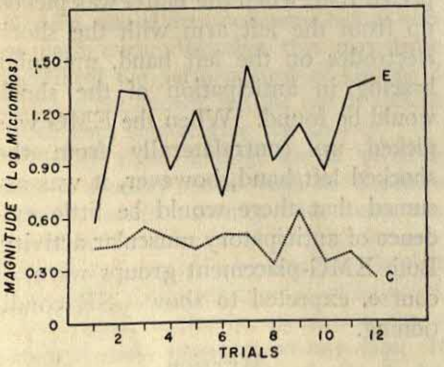


FIG. 1. Comparison of mean GSR magnitude during acquisition of paired (E) and unpaired (C) groups.

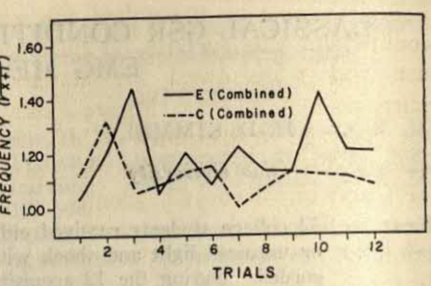


FIG. 2. Comparison of mean EMG frequency during acquisition of paired (E) and unpaired (C) groups.

RESULTS

The GSR dependent variable was 100 times the maximum change in log conductance in micromhos which occurred between 1.0 and 5.0 sec. from the onset of the CS. Figure 1 presents the mean GSR to the CS during training for the paired (E) and unpaired (C) groups, ignoring the location of the EMG electrodes. It is quite clear in Fig. 1 that the operation of pairing was effective in producing a difference in magnitude of the GSR between the two groups.[2] Analysis of variance of the GSR data showed that the difference between Groups E and C was significant, $F(1, 30) = 5.91$, Error $MS = .0847$. The interaction between conditions and trials did not approach significance.

The EMG dependent variable was the frequency of muscle action potentials (unintegrated) during the presence of the CS. Each pen deflection detectable by the naked eye was counted for this purpose. This measure was transformed for statistical purposes to $\sqrt{X+1}$. Figure 2 shows the average transformed EMG frequency during training for the paired (E) and unpaired (C) groups. It is clear in Fig. 2 that the pairing operation did not result in stable differences in mean EMG frequencies. Figure 3 presents the EMG data separately for the two paired groups, distinguished by whether the EMG electrodes were on the same side as the shock (E-S) or on the opposite side from the shock (E-D). A similar breakdown of the EMG data for the two unpaired groups (C-S and C-D) is shown in Fig. 4. Figures 3 and 4 show that regardless of whether pairing had been present, there was a slight tendency for an increase in EMG frequency to occur in the first two or three trials and that this increase tended to vanish thereafter. Neither location of the EMG electrodes appeared to result in conditioning of the EMG response.

---

[2] The arm containing the EMG electrodes is ignored in this comparison.

The actual independence of anticipatory GSRs and observed EMG changes is best illustrated by inspection of individual trials from selected *S*s' records. For this reason, a sample of such trials is presented to exemplify the different ways in which this independence may be manifested.[3] Figure 5, thus, presents photographs of the Grass recorder paper showing the fourth, sixth, seventh, and twelfth trials for one of the *S*s who received paired CS-UCS presentations and whose GSR and EMG were picked up from the same arm. The shock was delivered to the other arm. The figure also shows the record during the postexperimental instructed movements. On the fourth acquisition trial (Fig. 5a) this *S* made an unambiguous anticipatory GSR but showed no observable anticipatory EMG activity. Two trials later (Fig. 5b) both an anticipatory GSR and EMG activity are observable, yet, on the very next trial (Fig. 5b), only a GSR can be seen. On Trial 12 (Fig. 5c) the EMG record again indicated activity in conjunction with an anticipatory GSR. Figure 5d shows that even with considerable EMG activity in conjunction with instructed movements, this *S*'s GSR did not manifest discrete responses in any way

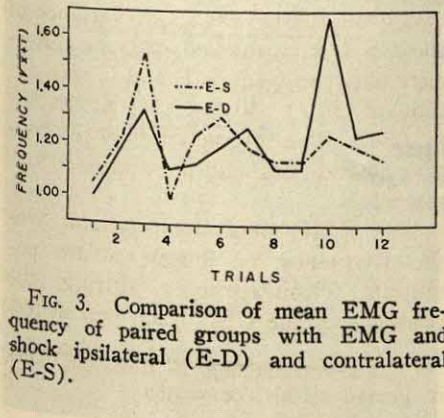


FIG. 3. Comparison of mean EMG frequency of paired groups with EMG and shock ipsilateral (E-D) and contralateral (E-S).

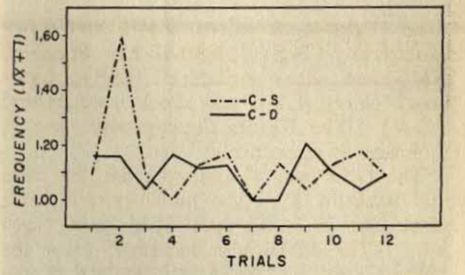
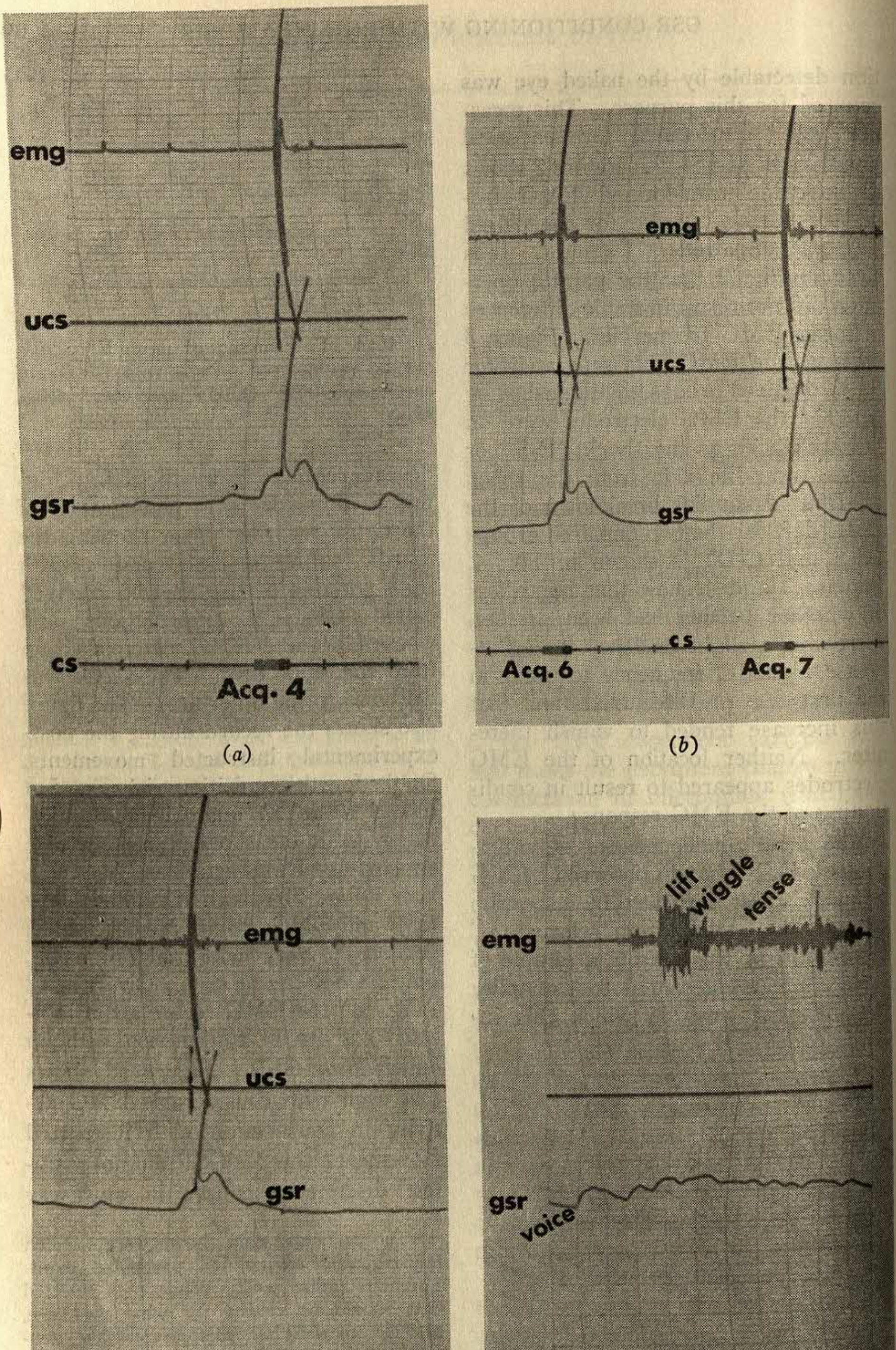


FIG. 4. Comparison of mean EMG frequency of unpaired groups with EMG and shock ipsilateral (C-D) and contralateral (C-S).

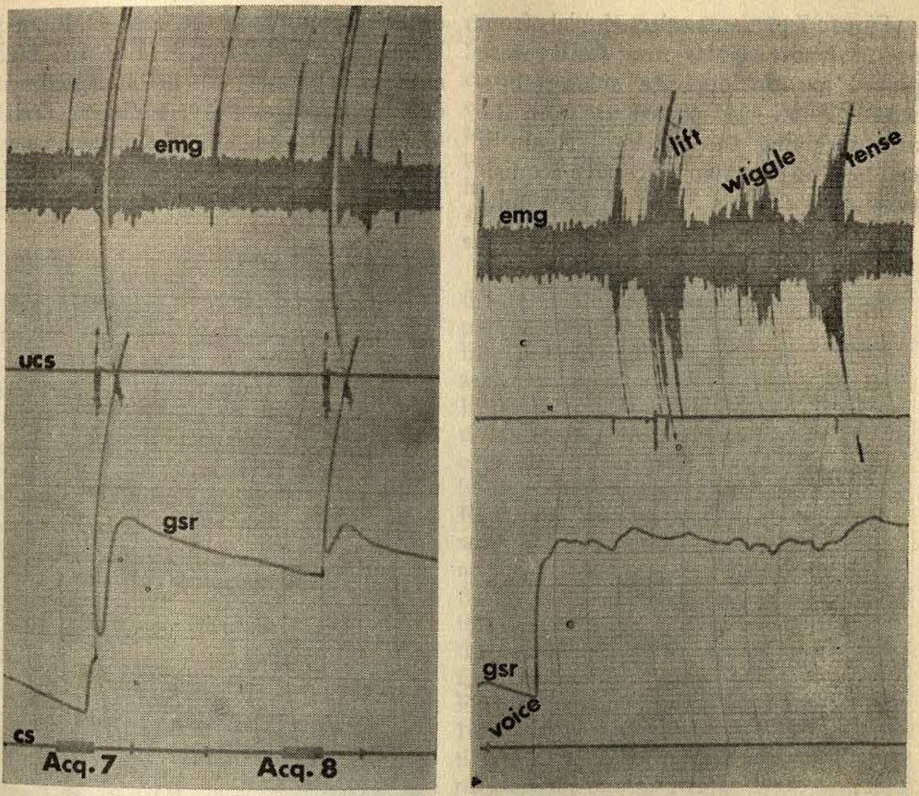
---

[3] The particular data shown were selected because they exemplified particular manifestations rather well. While such selected data should be viewed for what they are, *selected,* they do not misrepresent the general pattern of the present results. Rather, they are but a few of dozens of almost as good illustrations of similar manifestations of GSR-EMG independence.


emg
ucs
gsr
cs
Acq. 4
emg
ucs
gsr
cs
Acq. 6
Acq. 7
emg
ucs
gsr
cs
Acq. 12
lift
wiggle
tense
emg
gsr
voice

(a) (b)

(c) (d)



(*a*) (*b*)

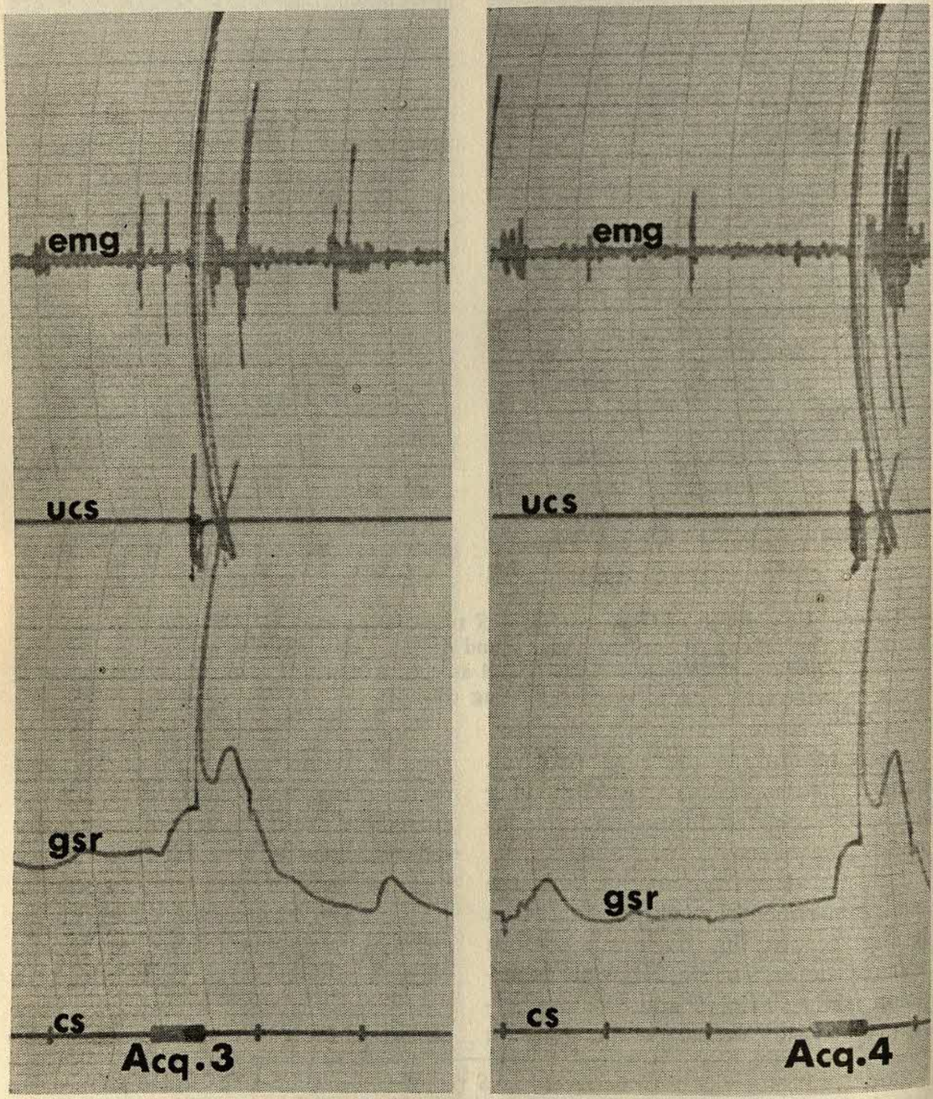
FIG. 6. Photograph of Grass record of *S* receiving paired CS-UCS with EMG contralateral to shock: (*a*) Acquisition Trials 7 and 8; (*b*) Postacquisition instructed movement. (Note the greater noise in this EMG record as compared with Fig. 5, which was collected under more favorable high-frequency filtering conditions.)

comparable to the observed CRs, although the drop in resistance following *E*'s voice persisted during the entire instructed sequence.

Selected data from another *S* in the same experimental condition are shown in Fig. 6. In Fig. 6a it can be seen that on the first of two adjacent acquisition trials (Trials 7 and 8) an anticipatory GSR occurred along with some EMG activity; but on the second of these two trials no GSR at all occurred but the EMG record was hardly different from before. Figure 6b shows that minimal GSR and substantial EMG activity were associated with instructed movement. Both Fig. 5d and 6b show that a large GSR was produced by *E*'s voice giving the *S* instructions to move.

FIG. 5. Photograph of Grass record of *S* receiving paired CS-UCS with EMG contralateral to shock. (*a*) Acquisition Trial 4; (*b*) Acquisition Trials 6 and 7; (*c*) Acquisition Trial 12; (*d*) Instructed movement at end of experiment. An electrical artifact appears in the GSR record following the UCS (produced by the operation of the UCS pen), making the distinction between the CR and the UCR quite clear. The GSR is represented by an upward pen movement.

Figure 7 presents selected trials from an *S* whose EMG and GSR were picked up on opposite sides (thus, whose EMG was picked up from the arm receiving the shock). Again it can be seen that two successive acquisition trials (Trials 3 and 4) give quite different pictures of the GSR-EMG relationship. On Trial 3 (Fig. 7a) the anticipatory GSR was accompanied by definite EMG activity, while on Trial 4 (Fig. 7b) a slightly larger and steeper GSR occurred without much EMG activity. Figure 7c presents the



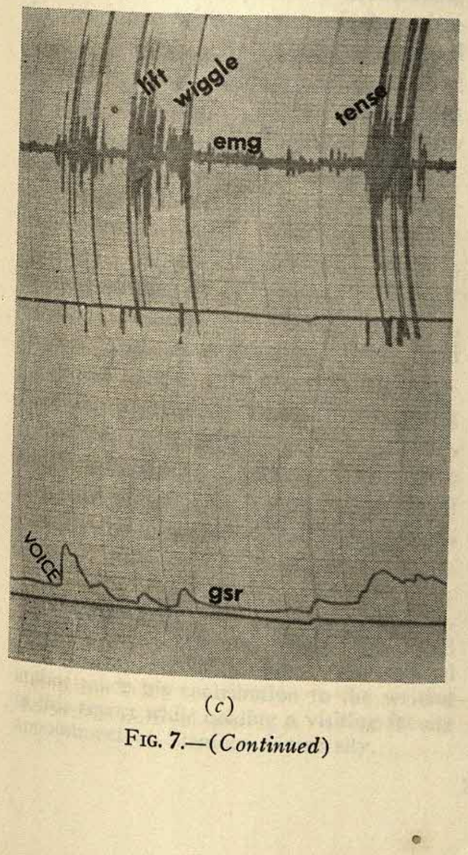
(*a*) (*b*)

FIG. 7. Photograph of Grass record of *S* receiving paired CS-UCS with EMG ipsilateral to shock. (*a*) Acquisition Trial 3; (*b*) Acquisition Trial 4; (*c*) Postacquisition instructed movement.

portion of this *S*'s data showing the EMG and GSR during instructed movement. Only the voice-produced GSR and the GSR accompanying tensing even approached the size of the anticipatory responses shown in Fig. 7a and 7b.

DISCUSSION

The results of this study indicate that the procedure of pairing a visual CS and a shock UCS, which is adequate to produce reliable GSR conditioning, did not result in EMG conditioning. In addition, inspection of trial-by-trial records of individual *S*s did not reveal any systematic tendency for conditioned GSRs to be associated with or preceded by EMG changes. With respect to the proposal by Smith (1954) that conditioning of the GSR is an artifact of muscular conditioning these results are decidedly negative. Not only was the GSR seen to occur in the absence of observable EMG changes, but there was not even evidence of EMG conditioning when the shock and EMG electrodes were on the same arm. If any tendency existed for *S* to "brace himself" to receive the shock, it was not observable in the present EMG records. Yet, instructed tensing of the arm *invariably* resulted in considerable EMG activity.



(*c*)

FIG. 7.—(*Continued*)

Apart from this study's contribution to the issue raised by Smith (1954), the present failure to obtain EMG conditioning is interesting in its own right. Even assuming independence of the GSR and EMG, it was expected that EMG conditioning would occur. Why was no EMG conditioning obtained? Inspection of the literature suggests two possibilities. First, and perhaps most obvious, the 5-sec. CS–UCS interval used in this study may have been too long for conditioning the EMG. Although there are no data on this point, it may be assumed that something closer to .5 sec. would be optimal for conditioning a muscular reaction. Furthermore, those studies (e.g., Fink, 1954) which have reported EMG conditioning, have used an instructed voluntary response as the UCR. It may be that electric shock fares no better as a UCS in conditioning muscular reactions in humans than it does with subhumans (Kimble, 1961).

The negative implications of these results with respect to the proposal that the GSR cannot be, itself, conditioned classically are limited to the muscular reactions recorded in this study. It is entirely possible that muscular reactions at other points in the body (e.g., trunk, neck), or those associated with respiration, were actually conditioned and were responsible for the GSRs observed. Taking Smith (1954) quite literally, we assumed that "bracing" to receive the shock would most likely be observed in the arm containing the hand that was shocked. As was indicated above, we assumed that such "bracing" might be found but that it might not be associated with observed GSRs. We further assumed that no such "bracing" would be seen in the hand contralateral to the shock, but that GSRs would be present in both cases. As it actually turned out, our expectations were not quite correct, but Smith's position hardly gained support thereby.

REFERENCES

Black, A. H., & Lang, W. M. Cardiac conditioning and skeletal responding in curarized dogs. *Psychol. Rev.*, 1964, **71**, 80–85.

Fink, J. B. Conditioning of muscle action potential increments accompanying an instructed movement. *J. exp. Psychol.*, 1954, **47**, 61–68.

Kimble, G. A. Hilgard and Marquis' *Conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Morrow, M. C. Recovery of conditioned UCR diminution following extinction. *J. exp. Psychol.*, 1967, **73**, in press.

Smith, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, **61**, 217–225.

Smith, K. Curare drugs and total paralysis. *Psychol. Rev.*, 1964, **74**, 77–79.

(Received April 19, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 75–83

STIMULUS SELECTION AT DIFFERENT STAGES OF PAIRED-ASSOCIATE LEARNING[1]

CARLTON T. JAMES AND JAMES G. GREENO

Indiana University

Ss learned paired associates with compound stimuli (word + nonsense). Transfer was tested to individual components after varying amounts of training on the compounds. Ss who were trained to a single perfect trial showed no more transfer to nonsense components than Ss trained to 4 of 8 correct responses to the compounds; however, Ss receiving 10 overtraining trials on compounds showed substantially more transfer to the nonsense components. Then a group of Ss was given training after learning a list of compound stimuli to criterion, but during the extended training other items were added so that the list was not mastered while the initial items were receiving overtraining. This group showed no more transfer to nonsense components than a group trained just to criterion. The data are generally consistent with the hypothesis that Ss actively select among stimulus aspects until the list is mastered, and then relax the selection of attention during overtraining.

It can be accepted now that when stimuli of paired associates are complex, *S*s often learn pairs by associating the response with a part of the stimulus. Underwood (1963) reviewed several items of evidence and concluded that *S*s carry on a process of stimulus selection, thereby producing a disparity between the nominal stimulus (presented by *E*) and the functional stimulus (with which *S* associates the response). Since Underwood's review, further studies by Newman and Taylor (1963), Jenkins and Bailey (1964), Cohen and Musgrave (1964), and Houston (1964) all have provided further demonstrations that after learning a list of pairs with complex stimuli, there are cases in which *S* can give the correct response to one element of a stimulus compound but not to another.

In the present experiments different groups were given different amounts of training on paired associates with compound stimuli. Then transfer lists were presented whose stimuli were the separate components from the original stimulus compounds. The response paired with a compound in the original list also was paired with both of its components in the transfer list. By measuring the amount of transfer to separate components after different amounts of training, information was obtained about the way in which stimulus selection occurs at different stages of learning.

EXPERIMENT I

Method

Materials.—It was desirable to have materials that would produce a strong selection effect. Items used were chosen from Cohen and Musgrave's (1964) lists. In the acquisition list, each stimulus was composed of a three-letter word and a CVC trigram. The words all had association values of 100%, and the average association value for the nonsense trigrams was 10% (Archer, 1960). The word always appeared on the left. The responses were the digits 1–8.

---

[1] This research was supported by the National Science Foundation under Grant GB 2791 to the second author. The second author made his contribution to the writing of this report while holding a visiting faculty appointment at Stanford University.

To construct the transfer list, each acquisition stimulus was split into its two components. Each component was paired with the same response as was its compound in the acquisition list. The transfer list contained all 16 paired associates formed in this way. Thus, each S received a transfer list containing eight pairs with word stimuli and eight with CVC nonsense stimuli.

Design.—Three groups of 20 Ss each were given different amounts of practice on the acquisition list. The Ss in the underlearning group (Group U) were stopped after the trial on which they first gave four correct responses. The Ss in the criterion group (Group C) were stopped after the trial on which they first gave correct responses to all eight of the items. And Ss in the overtraining group (Group O) were run to a criterion of 1 perfect trial and then run for 10 additional trials. Two Es each ran half of the Ss in each group.[2]

Procedure.—The acquisition list was presented using the RT or recall method; that is, study trials on the entire list were alternated with test trials on the entire list. There were 10 different orderings of the items, five study and five test. An equal number of Ss began the experiment on each of the five study randomizations. Since the correct responses were not shown on test trials, Ss in Groups U and C could learn no additional items after they first met their respective criteria of acquisition.

Study and test trials were presented on a memory drum at a 2-sec. rate, with a 2-sec. interval between trials. During acquisition, S was instructed to just study the items when the digit responses appeared, and to pronounce only the responses during the tests. The S was not forced to guess when he did not know the correct response. Instructions were identical for the three groups with one exception: S was told what criterion would be used for stopping practice on the acquisition list. The S was not told that there would be any test following the learning of the list.

Approximately 1 min. elapsed between the acquisition and the transfer lists. The transfer list was presented on flash cards with stimuli typed on one side and the correct response on the other. The stimulus was shown, and S was required to give a response to every item on every trial, including the first. There was no strict time limit on responding, but S was cautioned to respond within about 5 sec. in order to save time. After S responded, the correct response was shown for about 2 sec. The transfer list was presented for five anticipation trials.

Subjects.—The Ss were introductory psychology students at Indiana University, whose participation fulfilled a course requirement. One S was discarded for failure to follow instructions.

Results and Discussion

The mean numbers of trials taken by the groups to meet their respective criteria were 5.95, 9.45, and 22.45 for Groups U, C, and O. Comparability of the groups was checked by comparing performance on the acquisition list. Each S's score was the number of errors made through the first trial with at least four correct responses (the criterion trial for Group U). The 3×2 (Groups $\times$ Es) analysis of variance had no significant main effects. There was an inexplicable interaction between Groups and Es, $p < .01$, but this probably was not important since there was no significant effect due to different Es on the transfer list.

Two measures of performance on the transfer list will be presented. First, proportions of the word items and nonsense items which had zero errors in the transfer task will be reported. This gives an approximate measure of the proportions of components of each type for which S learned associations during acquisition. Secondly, the mean number of errors made by each S on each type of transfer item will be reported. The mean errors is a more conventional measure and probably permits more powerful tests of statistical significance.

The proportions of perfect items in the transfer task are presented in Table 1. The proportion of perfect items with word stimuli is quite a bit larger for Group C than for Group U, but it is nearly as large for Group C as for Group O. This is not surprising.

---

[2] We are grateful to Stanley Finger for assistance in running Ss.

TABLE 1

PROPORTIONS OF ITEMS WITH ZERO ERRORS IN TRANSFER

Group	Exp. I		Exp. II		Exp. III	
	Words	Trigrams	Words	Trigrams	Words	Trigrams
U	.287	.056	.356	.094	—	—
C	.600	.056	.638	.119	.740	.180
O	.681	.250	.838	.450	.930	.440
CE	—	—	—	—	.800	.194
OE	—	—	—	—	.920	.350

The significant result involves the proportions of perfect items with nonsense stimuli. Here there is a sizable difference between Group C and Group O, but the proportion of perfect items for Group C is small and equal to the proportion for Group U. This result suggests that very few associations were formed to nonsense components from the beginning of training to the time when the list was half learned. And apparently on the trials between half learning and criterion no associations were formed to nonsense components. However, after the criterion for the list was met, a substantial number of associations to the nonsense components were formed.

The same conclusions are indicated by the data in the form of total errors. Each *S* was given a score equal to the number of errors he made during the five transfer trials. For pairs with word stimuli the means of the error scores in Groups U, C, and O were 15.4, 7.8, and 5.5. The 3 × 2 (Groups × *E*s) analysis of variance yielded a significant *F* for difference among groups, $F\ (2, 54) = 9.72, p < .01$. The effect due to *E*s and the interaction between *E*s and groups were not significant. Duncan's test was applied to pairwise comparisons between groups. Group U differed significantly from both Group C and Group O, $p < .05$, but Group C and Group O did not differ significantly from each other.

For pairs with nonsense stimuli the means of the error scores for Groups U, C, and O were 27.2, 24.6, and 17.4. Analysis of variance yielded a significant *F* for variance due to groups, $F\ (2, 54) = 6.81$, $p < .01$. The effect due to *E*s and the Groups × *E*s interaction were not significant. By Duncan's test the difference between Group U and Group C was not significant, but Group O differed significantly from both Group U and Group C, $p < .01$.

EXPERIMENT II

The strong selection effect and the lack of difference between Groups U and C in Exp. I could be due to the confounding between position and meaningfulness in the acquisition list. In Exp. II, the physical location of the preferred component was not the same for every item.

Method

Materials.—The same words and nonsense syllables were used as in Exp. I, but the structure of the lists was changed. In four of the eight acquisition pairs the word appeared on the left and in the other four the word appeared on the right. Two lists were constructed; the pairs with the word on the left in one list had the word on the right in the other list.

A single transfer list was constructed. The list was obtained as in Exp. I, by dividing each of the eight acquisition items, thus producing a 16-item paired-associate list.

Design.—There were three groups of 20 *S*s each. All groups received a pretraining task to be described below. Then different amounts of practice were given on the acquisition list. The three groups (U, C, and O) were run to exactly the same criteria as their counterparts in Exp. I. Ten *S*s in each group were run on each of the two lists. The *S*s were run using a randomized block design. Each block consisted of six *S*s—one in each group with each list. There were three *E*s;[3] all *S*s in each block were run by a single *E*.

Procedure.—This experiment was run in the same way as Exp. I, with four exceptions. First, *S*s were given a pretraining paired-associate list. Each pretraining stimulus was composed of one of the letters B, S, T, F, K, and M, and one of the symbols #, $, %, &, *, and Σ. Three of the stimuli had the letter on the left and the other three had the symbol on the left. Responses were the digits 1–6. The *S*s were told that the main list would have the same structure as the pretraining list, but the stimuli would be compounds of words and syllables rather than letters and symbols. In the pretraining task, the items were presented on flash cards.

Secondly, we omitted any mention of the learning criterion from the instructions. This made the instructions identical for Groups U, C, and O.

Thirdly, in the instructions for the acquisition list, *S*s were told that they should try to respond to every item during test trials.

Finally, the transfer list was presented either for 10 trials or until *S* had given all 16 responses correctly on 2 trials.

Subjects.—The *S*s were students from an introductory laboratory course, participating to fulfill a course requirement. These *S*s had all completed the introductory course from which *S*s were taken for Exp. I. One was rejected for failure to follow instructions, and one for failure to learn the acquisition list within 40 trials.

Results and Discussion

Mean numbers of trials to criterion for Groups U, C, and O were, respectively, 2.65, 7.10, and 15.75. Groups were compared in acquisition performance by analyzing errors made through the first trial with four correct responses. In the 3 × 2 (Groups × Lists) analysis with randomized blocks the effects of groups and the Groups × Lists interaction were nonsignificant. There was a significant *F* for difference between the two lists, $p < .01$.

Proportions of items with zero errors in transfer are shown in Table 1. All the proportions are somewhat higher than in Exp. I, but the relationships among the groups are unchanged. Again, the transfer performance to nonsense components showed virtually no difference between Groups U and C, but Group O showed a considerable increment. Apparently the consistent position of the word component in Exp. I was not what caused the failure to form more than a few associations to nonsense components during early trials and the failure to form any associations at all to nonsense components during the trials between half learning and criterion.

As in Exp. I, impressions formed with proportions of perfect items are confirmed by analysis of total errors. For the word items, the mean numbers of errors per *S* in Groups U, C, and O were 16.45, 7.15, and 2.15. The experiment was a 3 × 2 factorial with randomized blocks. The error terms for treatment effects were the mean squares for Block × Treatment interactions. The effect due to Groups was significant, $F\ (2, 18) = 10.83$; $p < .01$. The effects of Lists and the Groups × Lists interaction were both nonsignificant.

For the nonsense items, the mean numbers of errors in Groups U, C, and O were 27.65, 27.75, and 10.20, respectively. The only significant effect in the analysis of variance was that for Groups, $F\ (2, 18) = 15.82$, $p < .01$.

We looked for effects of position and were somewhat surprised to find that none seem to have occurred. In acquisition, the mean number of errors through the first trial with four correct was 8.4 for items with words on the left and 8.7 for items with words on the right. The mean number of errors in transfer for words which had been on the left in acquisition was 4.7; there were an average of 3.9 errors in transfer for words which had been on the right in acquisition. The corresponding means for errors in transfer on nonsense items were 11.1 and 10.8.

---

[3] We are grateful to Lois Loats and Juris Lasmanis for assistance in running *S*s.

EXPERIMENT III

The results of Exp. I and II support the idea that *S*s used restricted attention until learning occurred, and then relaxed the selective process during overtraining. This relaxation of selectivity could occur in two different ways. First, the effect could have operated on individual items as they became learned. That is, the process of forming an association to the nonsense component of an individual item could have begun as soon as that item became learned. Alternatively, the effect could have operated on the list as a whole. That is, the process of forming associations to the nonsense components could have been delayed until all the items in the list were known.

The data from Exp. I and II give some support to the list-effect interpretation, because of the near equality between Groups U and C. The *S*s in Group C learned some of the items soon enough to have at least a few overtraining trials before the list as a whole met criterion. However, the argument for a list effect cannot be made strongly on the basis of these data. In Group C, some items received no overtraining trials and most items received very few trials after the associations were formed. Thus if stimulus selection was relaxed for an individual item as soon as it was learned, the procedure for Group C did not permit very many opportunities for associations to form to the nonsense components. Experiment III was carried out as a stronger test of the effect of list mastery on stimulus selection.

Method

Materials.—New lists of words and nonsense trigrams were constructed, of the same general type as those used before. The items for one acquisition list were ZOF EGG—1, AIR YAV—2, JOB RIW—3, ICE VUB—4, WUQ BAY—5, LAW NIJ—6, FEP CAR—7, GEX HAT—8. The other acquisition list had items with the same components, but with their positions reversed. The words all have a Thorndike-Lorge (1944) frequency of A or AA. The nonsense syllables all have association values from 0% to 20%, as measured by Glaze (1928). Three intermediate lists were constructed, each of which included the above listed items plus six new items. The stimulus for each new item was a compound of a CVC word with Thorndike-Lorge frequency of A or AA, and a CVC syllable with Glaze association value of 0–20%. In the first intermediate list the new items had the responses 9–14; in the second intermediate list the new responses were 15–20; in the third intermediate list the new responses were 21–26. The transfer list was constructed, as in Exp. I and II, by dividing the eight original acquisition stimuli into their respective components, and pairing each component with the response which went with the compound in the acquisition list.

Design.—There were four groups of 20 *S*s each. Group C was run to a criterion of one perfect trial, and Group O was run to one perfect trial and then given 10 trials of overtraining, as in Exp. I and II. A third group, Group CE, was given extended practice after the first perfect trial, but this practice was on one or more of the intermediate lists. As soon as *S* achieved a perfect trial he was shifted to the first intermediate list. If *S* achieved a perfect trial on the intermediate list before the tenth trial of extended training he was shifted to the second intermediate list. If a perfect trial on the second intermediate list occurred before the tenth trial of extended training, *S* was shifted to the third intermediate list. All *S*s in Group CE received a total of 10 trials of extended training except one, who achieved a perfect trial on the third intermediate list on his seventh extended trial, and who was given the transfer list immediately. (The small bias introduced by having an *S* in Group CE with only seven trials past criterion seemed less serious than the alternative of replacing an *S* in Group CE because he turned out to have a very fast learning rate.)

Note that Group CE and Group O are comparable in that each received training beyond the criterion trial. However, Group CE's overtraining trials included additional unlearned items in the list. For Group O, the trials constituted overtraining on the list; for Group CE, the trials constituted overtraining for the original eight items, but not for the total list being presented.

A final group (Group OE) was run to a criterion of one perfect trial on the original 8-item list, then to a criterion of one perfect trial on one 14-item list. Then 10 additional trials were given on the 14-item list. This group was run because of the possibility that overtraining effects would be suppressed in Group CE due to the increased length of the list, rather than to the presence of unlearned items in the list. Ten *S*s in each group were run on each of the two acquisition lists.

Procedure.—The acquisition and transfer procedures were exactly like those of Exp. II, except for the intermediate lists given to Groups CE and OE. When *S*s in Groups CE and OE reached the criterion on the first list, they were told that they were to learn another list, and that they would already know some of the items. Then the intermediate list was presented immediately in a second memory drum which had the list already in position.

Subjects.—Eighty-five paid *S*s were obtained from a pool of volunteers who responded to an advertisement in the student newspaper. The *S*s were students from all levels of class standing at Indiana University. Five *S*s were rejected for failure to complete the task within 45 min.

Results and Discussion

The mean numbers of trials taken by the groups to meet a first list criterion of one perfect trial were 5.50, 7.70, 5.35, and 5.40 for Groups C, CE, O, and OE. Groups were checked for comparability of performance during acquisition. Each *S*'s score was the number of errors made before the criterion trial on the acquisition list. The 4 × 2 (Groups × Lists) analysis of variance had no significant effects.

Proportions of perfect items are presented in Table 1. For the word items, the higher proportions for Groups O, CE, and OE relative to Group C were to be expected. The criterion of a single perfect trial is not strict, and some of the items given correctly on that trial undoubtedly were not learned when the training was stopped for Group C. That Group CE did not perform at as high a level on word items as Groups O and OE has two possible explanations. First, the learning rate on items in a long list is probably lower than on items in a shorter list. Another possibility is that *S*s attended more closely to the novel items in the intermediate list(s).

For the nonsense items, the large increase in the proportion of perfect items from Group C to Group O corresponds to the results of Exp. I and II. The proportion of perfect items in Group CE is hardly greater than Group C, suggesting that few if any associations to nonsense components were learned during trials which constituted overlearning on individual items. This suggests that associations were formed to nonsense syllables during overtraining because *S* extended his attention after he mastered the list, rather than because the individual items were learned. This conclusion is not contraindicated by Group OE, where the proportion perfect was somewhat higher than Groups C and CE, though not as high as Group O. Apparently the lengthening of the list to 14 items did not prevent the acquisition of associations to nonsense components, if training was extended to include overtraining on the longer list.

Again, the conclusions are supported by statistical analysis based on total errors. For the items with word stimuli, the average number of errors per *S* in transfer were 3.20, 2.85, .75, and

.65 for Groups C, CE, O, and OE, respectively. Analysis of variance showed a significant effect due to Groups, F (3, 72) = 5.33, $p < .01$. Neither lists nor the Groups × Lists interaction was significant. Duncan's multiple-range test indicated significant differences between Group C and Group O, $p < .01$, Group C and Group OE, $p < .01$, Group CE and Group O, $p < .05$, and between Group CE and Group OE, $p < .05$.

For pairs with nonsense stimuli the error scores for Groups C, CE, O, and OE, were 18.85, 16.05, 10.60, and 12.35. Analysis of variance showed a significant effect due to groups, F (3, 72) = 3.48, $p < .05$, but not due to lists nor the Groups × Lists interaction. Duncan's multiple-range test indicated significant differences between Groups C and O, $p < .01$, and between Groups C and OE, $p < .05$. Regarding the difference between Group O and Group CE, a difference between means of 5.70 is required to reject the null hypothesis with $\alpha = .05$. The obtained difference was 5.45.

That Groups CE and O did not differ significantly by Duncan's test is mildly disruptive to the argument, though it need not be taken too seriously. This experiment was run to permit a choice between two hypotheses. One hypothesis was that stimulus selection is relaxed after all the items in the list are learned. This list-effect hypothesis implies that Groups C and CE should give equal transfer to nonsense items, and Group O should be better. The alternative hypothesis was that stimulus selection is relaxed for each individual item after it becomes learned. This item-effect hypothesis implies that transfer to nonsense items should be the same for Groups CE and O, and Group C should be worse. The data leave little doubt as to which of these hypotheses is more reasonable.

We were concerned that part of the difference between Groups O and CE might have been due to an artifact. That Group CE transferred less well to word items than Group O suggests that during the extended training, *S*s in Group CE learned fewer items than Group O, or became confused about some that they knew when the extended training started. It might be that the nearly significant difference between Groups CE and O in transfer to nonsense items was due to less overall learning by Group CE or to general confusion introduced by the extra items. Therefore, we analyzed the transfer performance on items for which there were no errors in the 10 trials following the criterion trial. Counting only items with perfect postcriterion performance, the mean number of errors per item in transfer to nonsense components was 1.81 for Group CE and 1.16 for Group O. The difference of .65 error per item is almost exactly the same as the difference when all items are included. (Recall that for all eight nonsense items the mean difference per *S* between these groups is 5.45. Then the overall average difference per item was 5.45/8 = .68.) It seems safe to conclude that the difference in performance between Groups CE and O was not the result of confusion or attenuated learning rates which could have affected individual items in Group CE.

As in Exp. II, the position of the word in the compound had little or no effect. In acquisition, the mean number of errors per item (before criterion) was 2.07 when the word was on the left and 2.11 when the word was on the right. In transfer, the mean numbers of errors were .9 and 1.0 for words which had been on the left and on the right, respectively, during acquisition. The corresponding means for nonsense items were 7.1 and 7.9.

GENERAL DISCUSSION

The most straightforward interpretation of these results is that *S*s attend to stimuli in a selective manner during learning, and then relax the selective mechanism after the list has become learned. Interpretations like this are associated with the "noncontinuity theory" of the learning process. An alternative interpretation of our results could be given with continuity assumptions.

Using continuity theory, it would be necessary to assume a fairly high threshold, and higher rates of learning (larger increments) for associations to words than to nonsense components of the stimulus. Then during the first several trials of learning, the accumulation of associative strength for the nonsense components would exceed the threshold only rarely. Only when training was continued a long time beyond criterion would the associative strength accumulated for the nonsense components be sufficient to permit consistent correct responding.

Two relevant aspects of the results, based on data already presented, are discordant with the stated continuity hypothesis.

The first relevant fact is the main result of Exp. III. If the continuity interpretation were correct, we should have expected the extended training given Group CE to have added increments of associative strength to the nonsense components of the initial items, and brought them to a level approximately equal to the nonsense components in Group O. The continuity hypothesis has the same difficulty explaining the result of Exp. III as does the item-effect hypothesis.

The second relevant fact is the near equality in total errors between Groups U and C in Exp. I and II, and between Groups C and CE in Exp. III. The noncontinuity interpretation leads us to expect the same experimental result whether we look at proportions of items with zero errors or the mean number of errors that occur. But according to the continuity hypothesis, there is subthreshold learning which might not influence the proportion of perfect items, but which should influence the number of errors that occur in transfer before the item's strength exceeds the threshold. In Exp. II, the mean numbers of errors on nonsense components were exactly equal for Groups U and C, and in Exp. I and III, the advantage for the groups with more training was so small as to be easily explained on grounds of generalized practice and (in Exp. I) the fact that Group C had fewer word items that had not been learned at the beginning of the transfer task.

A third question was considered in the data, and will be mentioned briefly. The strongest form of the noncontinuity theory is an assumption that learning occurs in an all-or-none fashion, with parameters constant over trials. This assumption leads to a large number of quantitative predictions (Bower, 1961). Several features of the data from acquisition lists were compared with these predictions and reasonably poor correspondence was found between the theoretical and empirical results. However, the nature of the discrepancies suggested that the fault lay in assuming that the parameters were constant on all trials.

In both Exp. I and II, *S*s in Group U showed some transfer to nonsense components though the amount was small. This suggests that *S* probably attended to the entire stimulus early in learning, and then attended more selectively as learning proceeded. This possibility, along with other factors involved in adjusting to the task, could produce a lower probability of learning on the first trial or two of the experiment. Therefore, we recalculated the theoretical predictions under the assumption that different parameters applied on Trial 1 of the experiment, using techniques developed by Greeno (1966). (An illustrative application is available due to Greeno and Scandura [1966], where the special status of the first test was produced by transfer of training.)

The result was to give a satisfactory fit to most aspects of the data.

In summary, the weight of evidence seems to favor the noncontinuity interpretation. Our data do not decisively rule out a continuity interpretation, but they appear on balance to favor the noncontinuity view.

If the noncontinuity interpretation of our data is accepted, there are interesting implications regarding the role of attention in learning and the process of overlearning. The results suggest that *S*s actively select stimulus components during the course of learning. It seems unlikely that the selection is the result of a limited capacity to take in stimulus components, or of differing perceptual saliency of the components. Hypotheses which assume that selection occurs only in the perceptual system would be inconsistent with the marked increases in transfer to nonsense components that occurred when overlearning was carried out. Thus, it seems most reasonable to suppose that the learning process is intrinsically selective—that is, the selection of distinctive stimulus aspects which are easily stored in and retrieved from memory is one of the things that goes on as a part of learning.

The data suggest also that one of the processes occurring during overlearning is the addition of stimulus aspects or components to those selected during initial acquisition. It cannot be said whether this increased specificity results from separate associations formed between the added components and the response or from a process of association among the components which unites the parts of the stimulus into an integrated pattern. Either process could produce the results obtained in these experiments. But either process could also plausibly explain the fact that overtraining facilitates long-term retention. If associations are lost or become inaccessible over time, then items for which more than one stimulus aspect are connected with the response will be less likely to be forgotten. And one also might expect better long-term retention of items whose stimuli are integrated patterns.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monog.*, 1960, **74**(10, Whole No. 497).

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

Cohen, J. C., & Musgrave, B. S. Effect of meaningfulness on cue selection in verbal paired-associate learning. *J. exp. Psychol.*, 1964, **68**, 284–291.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Greeno, J. G. Paired associate learning with short term retention: mathematical analysis and data regarding identification of parameters. Mimeographed paper, Stanford University, 1966.

Greeno, J. G., & Scandura, J. M. All-or-none transfer based on verbally mediated concepts. *J. math. Psychol.*, 1966, **3**, 388–411.

Houston, J. P. S-R stimulus selection and strength of R-S association. *J. exp. Psychol.*, 1964, **68**, 563–566.

Jenkins, J. J., & Bailey, V. B. Cue selection and mediated transfer in paired-associate learning. *J. exp. Psychol.*, 1964, **67**, 101–102.

Newman, S. E., & Taylor, L. R. Context effects in paired-associate learning as a function of element-sharing among stimulus terms. *J. verbal Learn. verbal Behav.*, 1963, **1**, 243–249.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University Press, 1944.

Underwood, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 33–47.

(Received May 17, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 84–86

DIRECT AND ASSOCIATIVE PRIMING IN ANAGRAM SOLVING[1]

ROGER L. DOMINOWSKI[2] AND BRUCE R. EKSTRAND[3]

De Paul University *Northwestern University*

Anagram problems were attempted after 5 presentations of a 10-word list; 3 different lists were used. Compared to a control condition with no list, positive transfer was produced by prior exposure of the solutions, or of words associatively related to the solutions, with greater facilitation produced by exposure of the solutions. When words unrelated to the solutions were presented, but instructions stated that associations existed, negative transfer occurred. The findings were interpreted in terms of differences in the availability of solutions and were related to studies of category sets in problem solving.

Safren (1962) demonstrated that anagram solving is facilitated when solutions are associatively related and belong to a category. It was suggested that associative connections converge the pool of words from which *S* attempts to produce the solution, thus facilitating performance. A similar explanation may be offered for the facilitation found when *S*s study a list containing the solutions prior to solving anagrams (Mendelsohn & Griswold, 1964).

In the present study, a comparison was made of the relative effects of prior exposure of the solutions (direct priming) and of prior exposure of words related to the solutions (associative priming). The attempt was made to demonstrate positive transfer from associative priming with category relationships eliminated. In addition, one group of *S*s was given prior exposure of words unrelated to the solutions, but instructed that associations existed (inappropriate priming). This might be expected to produce negative transfer by producing a converged pool of words that does not contain the solutions.

METHOD

Materials.—Three different priming lists were used. The Direct list contained the solutions to the anagrams, CHAIR, LIGHT, WHITE, ROUGH, WATER, TIGER, HOUSE, SHORT, NURSE, and SMOKE. The words in the Associative list were TABLE, DARK, BLACK, SMOOTH, RIVER, LION, COTTAGE, LONG, DOCTOR, and TOBACCO. According to the Russell and Jenkins (1954) norms, each word in this list elicited one of the solutions as its strongest associate. The words in the Inappropriate list were WISH, JUMP, SQUARE, HUNGRY, APPEAR, CITIZEN, LIFT, RELIGION, NUMBER, and HAND. None of these words was associatively related to any of the solutions, according to the Russell and Jenkins norms.

Anagrams were constructed from the words in the Direct list by using 10 different two-move letter orders (Dominowski, 1966). Letter orders were randomly assigned to words; each anagram was typed in capital letters on a 3 × 5 index card.

Procedure.—The anagrams were placed in a 10 × 10 Latin square to control for possible sequence effects. Each *S* was tested individually. The instructions indicated that the letters of an anagram could be rearranged to make one English word and that a maximum of 120 sec. was allowed for each problem. An example of a four-letter anagram was shown. Solutions were given orally by *S*, and times were recorded by means of a stopwatch. If S did not solve within the time limit, he was given the solution before the next problem was presented. This was the complete procedure for Ss in the Control condition.

---

[1] This study was supported in part by a Grant-In-Aid-Of-Research from the Society of the Sigma Xi to the first author. The authors are grateful to Carl P. Duncan for his critical comments and to Joseph Regan for collecting the data.

[2] Now at the University of Illinois at Chicago Circle.

[3] Now at the University of Colorado.

After receiving anagram instructions, Ss in the Direct, Associative, and Inappropriate conditions received five presentations of their particular list. Five different random orders of the words were presented at a 2-sec. rate, using a continuous list on a Lafayette memory drum.

In the Direct condition, Ss were told that the words to be presented were the solutions to the anagrams. In the Associative and Inappropriate conditions, Ss were told that each word to be presented was strongly associated with one of the solutions. An example (MAN presented, WOMAN the solution) was provided.

Subjects and design.—As a course requirement, 50 introductory psychology students from De Paul University served as Ss. In addition to the priming groups (Direct, Associative, Inappropriate) two control groups were run. A schedule of 10 blocks of the five groups was constructed, and Ss were assigned to the schedule in order of their appearance at the laboratory. In one control condition, S was allowed to use pencil and paper while solving anagrams. However, only 1 S used pencil and paper. For purposes of analysis, the data from this S were discarded, and a single Control group of 19 Ss was used.

RESULTS

Number of solutions.—Each S attempted 10 problems; for the four conditions, the mean numbers of solutions were: Direct, 9.60; Associative, 9.30; Control, 7.95; Inappropriate, 6.50. Since Bartlett's test indicated significant heterogeneity of variance, an arc sin transformation was performed prior to statistical analysis. With the assumption of homogeneity of variance met, Duncan's multiple-range test was applied to the transformed data, using the procedure for unequal n's (McGuigan, 1960, p. 185). This analysis indicated that all differences were significant ($p < .05$) except that between the Direct and Associative means.

TABLE 1

ANAGRAM SOLUTION TIMES (IN SEC.) FOR EACH CONDITION

	Direct	Associative	Control	Inappropriate
Mean	1.95	7.25	31.00	54.35
Median	2.0	5.5	19.0	39.8
Range	1–3	2–16	2–116	5–120

Note.—Based on individual medians.

Median solution time.—Using each S's median solution time as the entry, adjacent groups (Table 1) were compared, using the Mann-Whitney U test (Siegel, 1956). Solution times in the Direct group were significantly faster than those in the Associative group ($U = 7, p < .002$). The Associative group solved significantly faster than the Control group ($U = 33$, $p < .02$). The difference between the Control and Inappropriate groups was not significant ($U = 73$, $p > .10$).

Problem analyses.—Since the same problems had been attempted in all conditions, the data were analyzed with problems as the replicate, and the groups were compared using the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956). For 9 of the 10 problems, mean solution times were faster in the Direct condition than in the Associative condition ($T = 5$, $p = .02$). For all problems, mean solution times in the Associative condition were faster than those in the Control condition. Seven of the problems had faster mean solution times in the Control condition than in the Inappropriate condition ($T = 6, p < .05$).

Differences in difficulty among problems were very consistent over conditions. Calculations of Kendall's Tau (Siegel, 1956) indicated significant agreement between the order of difficulty in the Control condition and the Direct (.51), Associative (.89), and Inappropriate (.60) conditions.

The words in the Associative list varied in terms of strength of association to the solution word. As this association becomes stronger, associative priming should become more like direct priming. Kendall's Tau was calculated between associative strength (from the Russell & Jenkins norms) and the difference in mean solution times between the Direct and Associative conditions, yielding a value of −.24. While not significant, this is the expected relationship.

Discussion

The results indicate that both direct and associative priming produce positive transfer to anagram solving, with greater facilitation from direct priming. Positive transfer seems adequately explained in terms of increased availability, although Mendelsohn and Griswold (1964) have argued that availability is not sufficient. Availability should be "maximal" following presentation of the solutions themselves; with prior exposure of associated words, not all *S*s associate correctly to each presentation, and facilitation is not as great. The effectiveness of associative priming should vary with the strength of association between primer and solution. There was a suggestion of this relationship in the present study.

It appears that inappropriate priming may impair anagram solving by increasing the availability of interfering responses. Such negative transfer is the result of many studies of set in problem solving, as pointed out by Schulz (1960). A series of "set" problems makes a certain type of response more likely. When this is followed by a problem requiring a different type of response, solution is more difficult. A clear demonstration of interference was reported by Gardner and Runquist (1958).

The effects of priming may depend on instructions; in the present study, *S*s were fully informed of the relationship between list presentation and anagram solving. Although direct priming has been found to facilitate under incidental learning conditions (Mendelsohn & Griswold, 1964), in other situations, attempts to increase availability (including free association) without complete information have not always been successful (e.g., Duncan, 1961; Maltzman, Brooks, Bogartz, & Summers, 1958).

As a final point, Safren (1962) argued that "category sets" operate through interitem associations. Studies of category sets typically confound category relationships with interitem associations. Facilitation by associative priming indicates that interitem associations can operate in anagram solving in the absence of category relationships. To the authors' knowledge, the reverse has yet to be demonstrated.

REFERENCES

Dominowski, R. L. Anagram solving as a function of letter moves. *J. verbal Learn. verbal Behav.*, 1966, **5**, 107–111.

Duncan, C. P. Attempts to influence performance on an insight problem. *Psychol. Rep.*, 1961, **9**, 35–42.

Gardner, R. A., & Runquist, W. N. Acquisition and extinction of problem-solving set. *J. exp. Psychol.*, 1958, **55**, 274–277.

Maltzman, I., Brooks, L. O., Bogartz, W., & Summers, S. S. The facilitation of problem solving by prior exposure to uncommon responses. *J. exp. Psychol.*, 1958, **56**, 399–406.

McGuigan, F. J. *Experimental psychology: A methodological approach.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Mendelsohn, G. A., & Griswold, B. B. Differential use of incidental stimuli in problem solving as a function of creativity. *J. abnorm. soc. Psychol.*, 1964, **68**, 431–436.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

Safren, M. A. Associations, sets, and the solution of word problems. *J. exp. Psychol.*, 1962, **64**, 40–45.

Schulz, R. W. Problem solving behavior and transfer. *Harvard educ. Rev.*, 1960, **30**, 61–77.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

(Received May 18, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 87–92

EQUIDISTANCE JUDGMENTS IN THE VICINITY OF A BINOCULAR ILLUSION[1]

FRANK L. AGEE, JR.[2] AND WALTER C. GOGEL

Boeing Company, Seattle, Washington

University of California, Santa Barbara

Misleading size cues associated with a binocularly observed trapezoidal window produced an apparent depth orientation of the window which was different from its physical orientation. As expected, it was found that errors occurred in adjusting 2 other objects (discs) to apparent equidistance with each other in the presence of the window, and that the direction of the errors in apparent equidistance was related to the direction of the errors in the perceived slant of the trapezoidal window. It was less clear that errors in the judgment of apparent equidistance occurred when the orientation of the window and the separation of the discs were vertical rather than horizontal. Possible explanations for discrepancies between the magnitude of the perceptual errors associated with the trapezoidal window and those associated with the equidistance judgments were discussed.

Research has clearly indicated that misleading size cues can alter the apparent depth orientation of an object in spite of the presence of veridical cues of binocular disparity (Ames, 1951; Gogel, 1954; Gogel, 1956; Gogel & Harker, 1955; Ittelson, 1952, 1960; Zegers & Malone, 1964; Zegers & Murray, 1962). For example, if an Ames trapezoidal window (Ames, 1951) is oriented with its smaller end closer to *O*, it will usually appear as though this end were the more distant. A situation of this type can be called a binocular illusion in depth since it is a distortion by other cues (in this case the cue of relative size) of the perception which normally would occur from the cue of binocular disparity. It has been demonstrated (Gogel, 1954) that the perceptual consequences of a binocular illusion in depth are not confined to the stereoscopic display used in producing the binocular illusion. More specifically, it has been found that the perceived depth position of an extraneous object will differ depending upon the directional position of this object with respect to the illusion display. The depth position of the extraneous object will be most correctly perceived with respect to the portion of the illusion display with which it is directionally most adjacent and less correctly perceived with respect to portions of the illusion display from which it is directionally more displaced (Gogel, 1954, 1956).

A demonstration of the effect of a binocular depth illusion upon the perceived depth of extraneous objects is schematically illustrated in Fig. 1. In Fig. 1, two differently sized playing cards (A and B), and three discs (1, 2, and 3) are physically located at the same distance from *O* and, according to the stereoscopic (binocular disparity) cue, should appear at the same distance. However, as a result of the

---

[1] The authors wish to thank James C. Hester for his assistance in collecting the data of this study.

[2] The research for this study was conducted at the Civil Aeromedical Research Institute by the first author under the direction of the second author, and was submitted by the first author in partial fulfillment of the requirements for the MS degree at the University of Oklahoma.

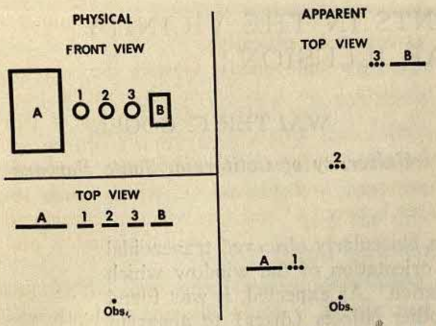


Fig. 1. Schematic drawing illustrating the effect of a binocular illusion in depth upon the apparent depth location of other objects.

misleading relative size cue between the two cards, Card A appears in front of (closer to *O* than) Card B. The difference between the physical and apparent distance positions of the cards in Fig. 1 represents a binocular illusion in depth. The discs in Fig. 1 are the extraneous objects, i.e., they contain no misleading size cues. The apparent depth positions of the discs in Fig. 1 approximate the mean results obtained in an experiment (Gogel, 1956) in which a single disc was alternately presented in each of the three lateral positions and judged as to its depth position with respect to each card. As is indicated in Fig. 1, Disc 1 correctly appears at the distance of the left card, but it incorrectly appears in front of the right card. Similarly, Disc 3 is seen correctly in depth only with respect to the playing card (Card B) with which it is most nearly in line of sight. Disc 2, which is directionally midway between the two cards, is correctly perceived with respect to neither card, but appears midway between them in depth.

It can be inferred from this and similar displays that if the discs had been adjusted to appear equidistant from *O*, Disc 3 would have had to be moved physically nearer to *O* than Disc 1, etc. (Gogel, 1954, p. 341). However, it is possible that in adjusting the discs to apparent equidistance with each other, *O* would be able to ignore the playing cards and consequently the effect of the binocular illusion upon the equidistance judgment would be reduced or eliminated. Also, as is indicated in Fig. 1, the effect of the binocular depth illusion extends beyond the immediate vicinity of the illusion-producing objects (playing cards). It is possible, therefore, that the magnitude of the distorting effect of the binocular illusion upon the apparent depth positions of the extraneous objects might vary as the extraneous objects are directionally displaced from the display producing the binocular illusion. Previous studies (Gogel, 1954, 1956) have employed stimulus displays in which the depth illusion has been oriented horizontally with the extraneous object separated laterally within (or nearly within) the right-left extent of the illusion. However, there is little reason not to expect similar results when the binocular illusion is vertically oriented with a vertical separation between the extraneous objects and when the extraneous objects are located outside the horizontal or vertical extent of the illusion. These conditions were investigated in the present study.

Method

Apparatus

Binocular illusion.—The binocular illusion in depth was produced by an Ames trapezoidal window (Portable Model, Catalog No. RTW_1), 33.8 cm. long, which was illuminated by a projector invisible to *O*. The average luminance of the light gray portions of the window was .35 ftl. Both horizontal and vertical orientations of the long axis of the window were used. The window was slanted in depth 45° with respect to *O*'s objective frontal plane when presented horizontally, but it was parallel to *O*'s objective frontal plane when presented vertically. The window presentations were such as to result in the following four experimental conditions: Cond. 1. A horizontal orientation of the window in which the large end of the window was on *O*'s left and was physically more distant than the small end. Cond. 2. A horizontal orientation of the window in which the large end of the window was on *O*'s right and was physically more distant than the small end. Cond. 3. A vertical orientation of the window in which the large end of the window was higher in *O*'s field of view than the small end, with the two ends physically equidistant from *O*. Cond. 4. A vertical orientation of the window in which the large end of the window was lower in *O*'s field of view than the small end, with the two ends physically equidistant from *O*. In all experimental conditions the large end of the window was 304 cm. from the eyes of *O*. Any effect of aniseikonia upon the overall experimental results for the horizontal orientations of the window was eliminated by orienting the window at equal but opposite directions of physical slant (Cond. 1 and 2). Any tendency to induce cyclorotation of the eyes when the window was vertically positioned was avoided by always placing the vertical window in a plane parallel to the objective frontal plane of *O* (Cond. 3 and 4).

Extraneous objects.—The extraneous objects which were adjusted by *O* to be at the same apparent distance (from himself) in the presence of the binocular illusion, were two circular spots (discs) of light with diameters of ½ in. The discs were slightly orange in color and were produced by Plexiglas surfaces transilluminated by incandescent lamps. By means of spur and rack gear arrangements, *E*, while invisible to *O*, could adjust the discs through approximately 60 cm. of depth. The luminance of the discs was .35 ftl.

One disc of the pair being judged remained stationary with its center always physically separated 4 cm. from the large end of the window and at the distance of that end (304 cm. from *O*). The path of depth movement of the adjustable disc was always perpendicular to the objective frontal plane of *O*, and for Cond. 1 and 2 was displaced laterally either 4 or 14 cm. from the small end of the window. The discs in Cond. 1 and 2 were at the height of *O*'s eyes (slightly below the midline of the window) and were laterally separated from each other by either 31.9 or 41.9 cm. For Cond. 3 and 4, only one vertical separation of the discs was used. The center of each disc was vertically separated 4 cm. from the end of the window most adjacent to it. Thus, a constant vertical separation of 41.8 cm. occurred between the discs when the window was oriented vertically.

The *O*s binocularly viewed the window and discs from a darkened observation position which contained a chin cup and a pair of eye pieces whose separation could be adjusted for different interpupillary distances. Apertures, invisible to *O*, which were located between *O* and the window, eliminated extraneous light so that only the window and discs were visible.

Slant measurement apparatus.—The occurrence of a binocular illusion in depth was a requirement of this study. Hence, in order to be certain that the illusion was present, the perceived slant of the window was measured. For this purpose, after every equidistance judgment made by *O*, a pair of lights was turned on in an area on *O*'s immediate left. This area contained a black rotatable indicator rod (¼ × ½ × 12 in.) and a gray reference board (to provide a physically frontoparallel plane) mounted on a light gray surface. The task of *O* was to adjust the indicator until it seemed to be slanted with respect to the reference board by the same amount that the window had appeared slanted with respect to his frontal plane. The indicator was horizontal or vertical as required by the horizontal or vertical orientation of the window. A scale, not seen by *O*, enabled *E* to record *O*'s slant adjustments in degrees.

Procedure

The *O*s were 16 men with a stereoacuity of at least 25.25 sec. of arc as determined by the Diagnostic Series for Macular Stereopsis of the Keystone Orthoscope. All *O*s made four consecutive equidistance judgments for each disc pair in each experimental condition. The sequence of stimulus presentations was systematically varied between *O*s.

For each adjustment, *O* used a small switch near the viewing position to signal *E* to move the adjustable disc in a particular depth direction until the movable disc appeared to be at the distance of (equidistant with) the stationary disc. After completing each disc adjustment, *O* was instructed to note the apparent orientation of the window, to face toward the area to his left which contained the slant apparatus, and to indicate the amount by which the window appeared slanted.

RESULTS

A summary of both the physical depth distance between the two discs

TABLE 1

*M*s AND *SD*s OF THE APPARENT SLANT OF THE WINDOW AND THE PHYSICAL POSITIONS AT WHICH THE DISCS APPEARED EQUIDISTANT

Directional Position of Small End	Horizontal Orientation of Window							
	Adjusted Depth Between Discs (in cm.)				Apparent Slant (in deg.)		Physical Slant (in deg.)	Amount of Window Illusion (in deg.)
	Small Lat. Separation		Large Lat. Separation					
	M	*SD*	*M*	*SD*	*M*	*SD*		
Small End Left	−25.8	4.8	−23.3	6.9	+45.6	8.3	−45.0	+90.6
Small End Right	−26.9	5.2	−25.7	8.9	+40.6	8.9	−45.0	+85.6
	Vertical Orientation of Window							
	M		*SD*		*M*	*SD*		
Small End Top	−4.1		4.6		+51.7	7.5	0	+51.7
Small End Bottom	+0.1		3.0		+35.0	8.6	0	+35.0

judged to be equidistant and the apparent slant of the trapezoidal window as measured by *O*'s adjustments of the rotatable indicator is given in Table 1. The values of apparent slant for the horizontal orientation of the window are averages of 32 scores, two from each *O*, where each score is an average of the four slant adjustments made with a particular orientation of the window and a particular lateral separation of the discs. The values of apparent slant for the vertical orientation of the window are averages of 16 scores, one from each *O*, where each score is the average of the four slant adjustments. A positive angle (slant) in the "Apparent Slant" column of Table 1 indicates that the small end of the window was perceptually more distant than the large end. A positive angle (slant) in the "Physical Slant" column of Table 1 would indicate that the small end of the window was physically more distant than the large end. Since the small end of the window was perceptually behind the large end for all orientations of the window, all the values in the "Apparent Slant" columns are positive. Since the small end of the window was physically in front of the large end for the horizontal orientations of the window and physically at the distance of the large end for the vertical orientations of the window, the "Physical Slant" values are negative or zero, respectively. The magnitude of the binocular illusion in depth is found by subtracting the physical from the apparent slant.

The average physical depth between the two discs when they appeared to be equidistant for the different conditions of the experiment also is shown in Table 1. It will be recalled that the movable disc was always directionally most adjacent to the small end of the window. A negative value in the adjusted depth between the discs indicates that, when the two discs appeared equidistant, the movable disc was physically closer to *O* than the stationary disc. In each horizontal condition, the disc pair, in order to appear equidistant, had to be physically separated in depth by amounts significantly different from zero beyond the .01 level of confidence with *t* (15) values of 20.20, 11.12, 20.67, and 13.14 for the values of −25.8, −23.3, −26.9, and −25.7 cm., respectively, of Table 1. Clearly, when the orientation of the window and the separation of the discs were horizontal, a large error occurred in adjusting the discs to apparent equidistance with each other. This error in adjustment was much less, however, when the orientation of the window and the separation of the discs were vertical. With the vertical condition, only the 4.1 cm. difference in depth adjustment was significant *t* (15) = 3.46, $p < .01$. The amount of distortion of the equidistance adjustments of the discs with the horizontal orientation of the window was only slightly different between the 31.9 and 41.9 cm. separations, with neither of these differences significant at the .05 level, *t* (15) = 1.96 and 0.87.

DISCUSSION

The results from the present study demonstrate that the presence of a binocular illusion in depth can affect the perceived depth between other objects even though these objects are extraneous to the illusion, and even though no portion of the binocular depth illusion (the window in this case) is directly involved in the task. It is as though the perceived space surrounding the (horizontal) window is distorted as a consequence of the binocular window illusion. For example, the disc directionally near the physically closer, but apparently farther, small end of the window had to be adjusted closer to *O* than the other disc in order for the two discs to appear equidistant. It is as though the apparent depth between the discs was distorted in the same direction as that of the ends of the window, and therefore, a depth adjustment opposite to this distortion had to be made in order for the discs to appear equidistant. The differences in the signs between the column labeled "Adjusted Depth Between Discs" and "Amount of Window Illusion" in Table 1 reflect this process.

Although the physical differences in depth between the apparently equidistant discs in Table 1 are large with the horizontal orientation of the window, there is reason to consider that this difference is not as large as would be expected from the magnitude of the depth distortion between the ends of the window. Physically the small end of the horizontal window was approximately 24 cm. in front of the large end. The stationary disc was always at the distance of the large end of the window, and, according to Table 1, the disc directionally near the small end of the horizontal window had to be adjusted approximately 25 cm. in front of the other disc in order for the two discs to appear equidistant. It follows that when the two discs appeared equidistant, they were physically located in depth near the ends of the apparently slanted window. This result is in agreement in direction but not necessarily in magnitude with the simplified hypothesis (Gogel, 1956, p. 1) that the binocular disparity between each disc and its directionally most adjacent end of the window will determine the apparent position of each disc.

There are several possible reasons why the slant illusion which occurred with the window might be expected to be greater than that between the two discs. First, the Ames trapezoidal window contains two (not one) binocular depth illusions. One of these is produced by the trapezoidal shape of the window and is measured by the *O*'s adjustment of the apparent slant of the window. The second binocular illusion is produced by the shading at the large end of the window and is used to simulate thickness in the window. The directions of these two illusions are *opposite* in their effect upon the error occurring in the equidistance adjustments of the discs. The presence of the second illusion would act to reduce the magnitude of the error in apparent equidistance from that expected from the first illusion. Another explanation can be found in what is called the "adjacency principle" (Gogel, 1963b). As applied to the present experiment, this principle states that the apparent depth position of each disc would be most determined by the binocular disparity cue which occurred between it and adjacent parts of the window. Binocular disparity between the disc and parts of the window displaced from the disc while not necessarily inconsequential would have less effect than adjacent parts in determining the apparent depth position of the disc. If the adjacency is defined as perceptual rather than as retinal or physical (Gogel, 1964), the magnitude of the error in the equidistance judgments would be expected to be less than the magnitude of the depth illusion occurring with respect to the window. A third possible explanation involves the task set of *O*. It is possible (Gogel, 1963a) that, to some extent at least, the act of performing the apparent equidistance adjustment of the discs permitted *O* to ignore the window, and thus to reduce somewhat its error-producing effects.

From Table 1, it is clear that the magnitude of the errors in the apparent equidistance judgments of the discs was less for the vertical than for the horizontal orientation of the window. This difference can also be explained by the opposing effects of the two illusions (size and shading) in the window. The magnitude of the first illusion clearly increases with an increase in the physical slant of the window (see Table 1) while the magnitude of the shading illusion does not. It follows that the capacity of the second illusion to offset the effect of the first upon the apparent equidistance adjustments of the discs should be much greater for the frontoparallel (vertical) rather than the 45° (horizontal) orientation of the window. It is likely that, with the vertical window, the resultant effect of the two depth illusions was small, and that, therefore, the error found in the equidistance judgments also was small.

The increase in the horizontal separation of the discs (from 31.9 to 41.9 cm.) with the window oriented horizontally did not produce a large change in the adjustment error (Table 1). The results suggest that the effect (*a*) extends beyond the edges of the window, but (*b*) with slightly decreasing magnitude. The last portion of this conclusion must be considered as very tentative pending the use of additional separations.

REFERENCES

Ames, A., Jr. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65** (7, Whole No. 324).

Gogel, W. C. Perception of the relative distance position of objects as a function of other objects in the field. *J. exp. Psychol.*, 1954, **47**, 335–342.

Gogel, W. C. Relative visual direction as a factor in relative distance perceptions. *Psychol. Monogr.*, 1956, **70** (11, Whole No. 418).

Gogel, W. C. Size cues and the adjacency principle. 1963, *FAA Civil Aeromed. Res. Instit.*, Rep. No. 63-28. (a).

Gogel, W. C. The visual perception of size and distance. *Vision Res.*, 1963, **3**, 101–120. (b)

Gogel, W. C. Perception of depth from binocular disparity. *J. exp. Psychol.*, 1964, **67**, 379–386.

Gogel, W. C., & Harker, G. S. The effectiveness of size cues to relative distance as a function of lateral visual separation. *J. exp. Psychol.*, 1955, **50**, 309–315.

Ittelson, W. H. *The Ames demonstrations in perception.* Princeton: Princeton Univer. Press, 1952.

Ittelson, W. H. *Visual space perception.* New York: Springer, 1960.

Zegers, R. T., & Malone, T. B. Perception of distortion. An experimental study of distance and perspective in the Ames trapezoidal illusion. 1964 USN Training Device Center, Tech. Rep. No. NAVTRADEVCEN 1215-1.

Zegers, R. T., & Murray, A. Perception of distortion. I. An experimental approach to illusion. 1962 USN Training Device Center, Tech. Rep. No. NAVTRADEVCEN 506-1.

(Received May 21, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 93–98

SELF-REINFORCEMENT AND EXTERNAL REINFORCEMENT IN VISUAL-MOTOR LEARNING

ALBERT MARSTON

University of Wisconsin

2 experiments examined the effect, upon learning, of reinforcement introduced between a motor or perceptual response and subsequent information feedback. In Experiment I, blindfolded Ss threw darts and were told where on the target a dart had landed. 3 reinforcement treatments varied in the events intervening between S's response and the feedback: (a) external reinforcement (ER), E gave poker chips for bull's eyes; (b) self-reinforcement (SR), S instructed to take chip when he judged himself correct; (c) no reinforcement in the interval. Contrary to prediction, ER showed the overall poorest performance. In Experiment II, S judged the length of lines. In addition to ER, SR, and information-only (IO) conditions, varying as in Experiment I, there were 3 levels of amount of information feedback. It was found again that ER produced poorest performance in early trials. The groups receiving no information did more poorly than either of the other information conditions, which did not differ significantly from each other. There was no interaction of reinforcement and amount of information.

A number of recent studies have investigated self-evaluative responses in a learning situation (Kanfer & Marston, 1963a, 1963b; Marston, 1964, 1965; Marston & Cohen, 1966; Marston & Kanfer, 1963). While a number of the variables in these experiments have been concerned with determinants of positive self-evaluation, the structuring of the evaluation response as the self-administering of freely available rewards (points, poker chips, prizes) has raised the question of the reinforcing function of such positive self-evaluations. Marston and Kanfer (1963) approached this question by examining postacquisition performance under conditions of extinction (i.e., absence of external rewards) vs. extinction plus the opportunity to administer self-reinforcement (SR). They found, in the task of memorizing the correct nonsense syllable among sets of four, that without SR performance declined over postacquisition trials and that with SR criterion level (70% correct) performance was maintained. Subsequent studies examined SR in this general paradigm, acquisition followed by SR trials without external feedback. Marston (1964) was unable to replicate the extinction (or forgetting) phenomenon in the postacquisition trials of the nonsense-syllable task, thus raising a question about the reinforcing effect of positive self-evaluations (self-administered rewards—SR).

The purpose of the present study was to examine the relative reinforcing effects of SR and externally controlled reinforcement *during* acquisition. The *S* made a response which was followed, after a brief interval, by feedback concerning response accuracy. In that interval, for three separate experimental groups, three types of events could occur: (*a*) external reward (poker chips) if the response were accurate, (*b*) self-reward (taking of poker chips) if *S* judged his response as accurate, (*c*) nothing. Unlike the previous studies on SR, the present one utilized visual tasks constructed so that accuracy was ambiguous enough to make *S* dependent on the subsequent, explicit external feedback to improve his performance. Experiment I used dart throwing with *S* blindfolded. Experiment II used judgment of length of lines and also varied the quality of the external feedback. In both experiments the primary hypothesis was that reinforcement would produce faster learning than information alone, with externally controlled reinforcement superior to self-reinforcement.

EXPERIMENT I

Subjects.—Fifty-nine right-handed, undergraduate males volunteered as *S*s in order to earn extra credits toward their grade in introductory psychology at the University of Wisconsin.

Apparatus.—The *S* was seated at a table with his right elbow resting on a foam pad near the edge of the table. On the wall opposite him was placed a wooden dart board, 17 in. in diameter. The dart board was 47 in. from the elbow pad and 34.5 in. from the floor (from the bottom edge of the board). A 6-in. bull's-eye was painted on the board. The remainder of the target was divided into four equal quadrants by a horizontal line and a vertical line painted through the diameter. The *S* wore a blindfold during the task and threw steel pointed darts at the target. Just to the left of *S* on the table was placed a tubular container of poker chips. A metal, sliding bar, when pulled, released one chip into a box hanging below the edge of the table.

Procedure.—All *S*s were told that the study involved the investigation of motor skills used in target tracking under minimal visual stimulation. Each was given one practice shot without the blindfold and told to keep his elbow on the foam pad and to try to keep his hand steady. It was also indicated that for all trials with the blindfold *E*[1] would tell *S* whether he hit the bull's-eye, and if not, which of the four quadrants he hit. The *S* was assured that *E* was interested only in motor learning and that all feedback information would be completely accurate. He was told that *E* would place the dart in his hand at the start of each trial. The *E* stood to the left and slightly behind *S*.

Three treatment groups varied only in the events taking place between the dart-throwing response and *E*'s verbal feedback for each of the 60 trials. For the Information Only (IO) group, *S*s received only the verbal feedback as to where each dart landed and were given no further instructions. For the External Reinforcements (ER) group, for those trials on which *S* hit the bull's-eye, *E* released a poker chip into the box before announcing, "That was a bull's-eye." For all other trials, *S* was told only where the dart landed. Additional instructions for this group and for the SR groups informed *S* that he was to try to earn as many poker chips as possible and that for each set of 15 chips he would receive additional credit for serving in the experiment. Two Self-Reinforcement (SR) groups were included; both receiving the information given IO but both also receiving instructions to take a poker chip after any trial on which they thought they had hit the bull's-eye. The two groups differed in one sentence of the instructions: one (SR Facil) was encouraged to be self-confident and take a chip even though uncertain about hitting the bull's-eye; the second (SR Inhib) was told to be cautious and take a chip "only when you are very confident."

The interval between dart-throwing response and verbal feedback was kept at a constant 5 sec. for all groups, with SR groups told they would be given 5 sec. to decide whether to take a chip.

The *S*s were assigned in sequence to the two SR, the ER, and the IO groups. Each of the four groups contained 15 *S*s. However, in the subsequent statistical analysis, errors revealed in the data recording for one *S* in the ER group necessitated exclusion of his data from the analyses.

Results and discussion.—An analysis of variance on the arc-sine transformation of proportion SR responses per block of 10 trials yielded no significant difference between the SR-Facil and SR-Inhib groups, $F < 1.00$. While similar instructions affected SR rate in an earlier study (Kanfer & Marston, 1963a), in which instructions were given just prior to a postacquisition test period, facilitation or inhibition was not demonstrated during acquisition in the present study. For both groups, proportion of SR increased over blocks, $F\ (5, 140) = 7.13$, $p < .01$, with proportion SR maintaining a level slightly above proportion of correct responses.

---

[1] C. Yosowitz and B. Stone served as *E*s.

Table 1 gives the mean proportion of correct responses (bull's-eyes) over trial blocks for the four treatment groups. The analysis of variance of these data showed a significant effect of reinforcement-treatment groups, $F\ (3, 55) = 2.81$, $p < .05$, and a significant interaction of Treatment and Trial Blocks, $F\ (15, 275) = 1.92$, $p < .05$. Using a Tukey gap test ($p < .05$) it was found that the ER group showed a significantly lower mean over blocks than any of the other groups, which did not differ significantly from one another.

A trend analysis revealed significant differences among the groups in linear component, $F\ (1, 275) = 5.42$, $p < .05$; quadratic component, $F\ (1, 275) = 6.44$, $p < .05$; and in cubic component, $F\ (1, 275) = 6.55$, $p < .05$. The analysis pointed to an initial superiority of IO over ER, with the former group showing an erratic and nonsignificant overall increase in accuracy. By the last trial block, ER and IO were not significantly different. It appears that in terms of initial performance, and even on overall means, external reinforcement reduces accuracy relative to information-only and information plus self-reinforcement. However, by late in the task, the effect of external reinforcement brought performance up to the levels of the IO and SR conditions.

An analysis of reinforcements for correct responses, whether SRs or ERs, yielded no differences among the three reinforcement treatment groups, $F < 1.00$. An analysis of overall number of reinforcements did show a significant Group effect; $F\ (2, 205) = 4.22$, $p < .05$; indicating that the SR groups took a significant number of self-reinforcements for incorrect responses.

TABLE 1

MEAN PROPORTION OF CORRECT RESPONSES (BULL'S-EYES) IN DART THROWING FOR THE FOUR REINFORCEMENT TREATMENTS

	Trial Blocks					
ER	.21	.30	.23	.42	.42	.47
SR-facil	.32	.35	.58	.51	.48	.49
SR-inhib	.30	.35	.56	.61	.59	.49
IO	.39	.49	.62	.49	.45	.56

An analysis of the repetition of particular types of errors (i.e., in specific quadrants) from trial to trial indicated no significant group differences; $F\ (3, 55) = 1.92$, $p > .10$. Rigidity of type of error decreased over trial blocks; $F\ (5, 269) = 4.27$, $p < .01$.

EXPERIMENT II

The second experiment replicated Exp. I, using a different learning task, namely judgment of length of lines. In addition, Exp. II, using a factorial design, examined the reinforcement variable at three levels of information feedback (exact length of line, "greater" or "lesser" than *S*'s response, no feedback), yielding nine experimental groups. It was predicted that there would be an interaction between quality of feedback and reinforcement, with the effects of Exp. I shown at high levels of feedback and the originally expected ER > SR > IO relationship shown at the no-feedback level.

Subjects.—The sample consisted of 162 male undergraduate volunteers from the introductory psychology course at the University of Wisconsin.

Apparatus.—The *S* was seated at a desk, separated from *E* by a 2-ft. sq. screen containing a window for presentation of stimulus cards. The *S* was approximately 2 ft. away from the stimuli. A tubular container at *S*'s left was used to dispense poker chips into a box.

The stimuli consisted of 10 4 × 6 in. white cards, each containing a $\frac{3}{16}$ in. wide, horizontal black line. The length of the lines ranged in $\frac{1}{8}$-in. steps from 1 in. to $2\frac{1}{8}$ in. The 10 cards were duplicated in 10 different random orders.

Procedure.—The *S* was told that the experiment dealt with perceptual skills relevant to target tracking in radar work. He was then instructed to estimate the length of each line, having been informed of the 10 possible lengths and told that the stimuli would be shown in random order (indicating that the same length could be repeated on adjacent trials).

The *S* was given instructions regarding the nature of his feedback and reinforcement, if the latter were involved. High information (HI) instructions stated that *E* would say "That's correct" or give the actual length if *S* were incorrect. Medium Information (MI) instructions were the same as HI, except noting that *E* would say "Longer" if the actual line were longer than *S*'s response or "Shorter" if the stimulus were smaller than *S*'s estimate. No Information (NI) *S*s were told only to give their responses, with no mention made of information feedback. Reinforcement conditions were described as in Exp. I, with ER groups told *E* would release a poker chip if *S* were correct, prior to giving information feedback (for ER-NI, *S* received only a chip when correct and no verbal feedback). The SR groups were told to take a poker chip whenever they felt "fairly confident" their estimate was correct. No mention of reward was made to the three IO groups. Thus, the condition Information Only, No Information (IO-NI) consisted simply of practice at the task with no feedback from *E* and no reward, either from *E* or in the form of SR.

Each *S* received 10 blocks of randomized presentations of the 10 stimulus cards. The *E* gave *S* 5 sec. to take a chip in the SR groups. In the ER groups, *E* released the chip immediately following each correct response. There was an interval of approximately 5 sec. between trials. Each stimulus was presented for 5 sec.

Results and discussion.—An analysis of variance on accuracy of performance was done on the arc sine of the proportion of absolute amount of error in size judgment divided by the size of the stimulus. Table 2 presents the means across blocks for the nine groups. There was a significant effect of information feedback, $F\ (2,\ 153) = 24.33$, $p < .01$, due to the much poorer performance of the NI groups. The MI and HI treatments did not differ significantly.

For all groups there was a significant decrease in relative error over trials, $F\ (9,\ 1{,}377) = 9.00$, $p < .01$. A trend analysis was performed to

TABLE 2

MEAN PROPORTIONS: ABSOLUTE ERROR IN LINE JUDGMENT
DIVIDED BY SIZE OF STIMULUS

Treatment	Trial Blocks									
	1	2	3	4	5	6	7	8	9	10
HI-ER	.04	.03	.03	.04	.04	.04	.04	.03	.03	.02
HI-SR	.05	.03	.04	.03	.04	.04	.03	.02	.03	.03
HI-IO	.04	.02	.03	.03	.04	.03	.03	.02	.03	.03
MI-ER	.05	.03	.04	.04	.04	.03	.03	.02	.04	.02
MI-SR	.05	.03	.03	.03	.04	.03	.03	.03	.04	.03
MI-IO	.03	.03	.02	.03	.03	.04	.03	.03	.03	.03
NI-ER	.06	.04	.04	.05	.05	.04	.05	.05	.05	.04
NI-SR	.05	.04	.04	.04	.05	.05	.05	.05	.06	.05
NI-IO	.06	.04	.05	.05	.05	.06	.05	.05	.06	.06

clarify the significant Reinforcement × Trials effect, F (18, 1,377) = 2.00, $p < .05$. The three reinforcement conditions all differed significantly from one another (Tukey gap test at $p < .05$) in linear trend (mean linear coefficients were IO = +.12, SR = −.25, ER = −.97). The ER and SR conditions both showed a negative slope (decrease in error), with IO showing a slight positive slope.

There was some evidence, as in Exp. I, of a decrement in performance for the ER condition. On Trial Block 1, ER showed the greatest amount of error, with a significant difference between ER and IO (Tukey gap test at $p < .05$). However, by Block 10, ER evidenced significantly less error than either SR or IO.

In addition to examining the effect of the treatment on accuracy, per se, it was possible to evaluate general variability in the *S*'s responses. Although all 10 stimulus lengths were presented in each block of trials, *S* was not informed of this, allowing him to repeat a given length within the trial block. While frequency of repetitions within a block would be correlated with accuracy, this fixation measure provides a possibly sensitive indicator of performance.

An analysis of variance of frequency of within trial-block repetitions indicated the significant interaction of Reinforcement × Blocks; F (18, 1,367) = 1.74, $p < .05$, and Information Feedback × Blocks; F (18, 1,367) = 1.58, $p < .05$.

At Block 1, ER had the greatest mean frequency of repetitions, with SR and IO showing significantly lower means (Tukey gap test, $p < .05$). While SR and ER evidenced significant Block 1–Block 10 decreases in repetitions, IO increased significantly. By Block 10, SR and ER displayed significantly lower mean frequency of repetitions than IO. (Block 10 means were IO = 3.17, SR = 2.67, ER = 2.87.)

Thus the analysis of the repetition measure supported the findings of the analysis on amount of error. External reinforcement produced an initial decrement, but resulted in a subsequent superiority in performance. SR lay between ER and IO on initial level of performance, and was equivalent to ER in superiority over IO at the final trial block.

Analysis of the arc-sine proportion SR per block of trials in the HI, MI, and NI treatments, yielded no significant group differences. Over all groups, frequency of SR increased significantly over trials, F (9, 459) = 5.4, $p < .01$.

CONCLUSIONS

These two experiments have fractionated the operation usually termed reinforcement in human learning research. Whether one refers to these two operations as information vs. reinforcement per se (or incentive), or one conceives of them as two types of generalized reinforcers, the present results point to a somewhat surprising effect of the reinforcing or incentive operations when they are administered by *E*. At least initially, on these visual-motor tasks, the addition of what operationally would be described as incentive to information alone debilitates rather than enhances performance.

One interpretation of the findings would be that whereas ER may be dissonant with the covert, self-evaluative responses likely to occur immediately following any skill performance, SR is much more likely to be consonant with the covert behavior. As the task progresses, *S* either suspends self-evaluation while anticipating ER, or his self-evaluation becomes more consonant with external evaluation. This interpretation assumes that the evaluative aspects of the subsequent information feedback do not have the same effect as ER, either because of the temporal delay, or because of the evaluative emphasis given ER in the instructions. Regardless of the interpretation of the present findings, it seems clear that the scaling of incentive values of human reinforcers must include the possibility that evaluations defined as positive by *E* can have in fact negative effects.

The same incentives self-administered seem to have a more uniformly positive effect than external reinforcements. Self-evaluations in the present study can be described as reinforcing events, in the simply operational sense of increasing the probability of correct response above that produced by information alone. In a theoretical context, it is not yet clear whether SR acts on habit strength or is a more general, incentive phenomenon directly affecting response probability.

REFERENCES

Kanfer, F. H., & Marston, A. R. Determinants of self-reinforcement in human learning. *J. exp. Psychol.*, 1963, **66**, 245–254. (a)

Kanfer, F. H., & Marston, A. R. Conditioning of self-reinforcing responses: An analogue to self-confidence training. *Psychol. Rep.*, 1963, **13**, 63–70. (b)

Marston, A. R. Response strength and self-reinforcement. *J. exp. Psychol.*, 1964, **68**, 537–540.

Marston, A. R. Imitation, self-reinforcement, and reinforcement of another person. *J. pers. soc. Psychol.*, 1965, 2, 255–261.

Marston, A. R., & Cohen, N. J. The relationship of negative self-reinforcement to frustration and intropunitiveness. *J. gen. Psychol.*, 1966, **74**, 237–243.

Marston, A. R., & Kanfer, F. H. Human reinforcement: Experimenter and subject controlled. *J. exp. Psychol.*, 1963, **66**, 91–94.

(Received May 11, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 99–105

SERIAL ORDER IN RECOGNITION AND RECALL[1]

M. C. CORBALLIS[2]

McGill University

2 experiments are reported in which Ss viewed 100 series of 5 digits, each followed by a 6th "critical" digit. They were required to indicate as quickly as possible by pressing 1 of 2 buttons whether or not the critical digit had appeared in the series. In Exp. I (20 Ss) presentation rate was varied. In Exp. II (10 Ss) there were 3 recall trials as well as the 100 recognition trials, but S did not know whether he was to recall the 5 digits or make a recognition response until after the digits had been presented. In both experiments recognition times showed a strong recency effect, i.e., faster recognition times the later the critical digit occurred in the series, while in the recall trials in Exp. II Ss generally reported the digits in the forward order. These results are interpreted to mean that the forward ordering in recall is a feature of the recall process itself rather than of storage, and that in recognition S has direct access to relevant elements of a serial trace without first having to scan irrelevant ones.

Wickelgren (1966) has distinguished between two kinds of models for short-term storage of sequentially presented items. One of them, described as a "non-associative" model, maintains that the items are encoded into an ordered set of storage locations. The ordering of these locations is fixed, so that *S must* "read" out the items in the order in which they were encoded. Milner (1961) has speculated about the neural bases of such a model, and Conrad (1965) has shown how order errors may occur through *S*'s attempts to reconstruct the rapidly fading information which is stored in each location.

The other kind of model, the one preferred by Wickelgren (1965, 1966), is described as an "associative" model. This model holds that short-term associations are established between the internal representatives of items presented in close temporal contiguity. Order of recall would then be determined by these associative connections.

Both of these models imply that the forward ordering typically found in short-term serial recall is a function of the way in which the items are stored. But the possibility also exists that the ordering of items in recall is a feature of the recall process itself, not of storage. The two experiments reported here were designed to test this possibility. Experiment I was a recognition experiment in which *S* was presented with a series of five digits followed by a sixth, "critical" digit. His task was to indicate as quickly as possible whether or not the critical digit was included among the previous five. This technique is similar to that reported by Norman and Wickelgren (1965) except that their study was concerned with recognition *accuracy* while the present study was mainly concerned with recognition *time*. If the forward ordering of items in recall is a function of *storage,* we should expect recognition times to show a primacy effect, but no recency effect; that

---

[1] This research was supported by a Contract, Nonr-4896(00), between the Office of Naval Research and McGill University and by a research grant (9425-10) from the Defence Research Board of Canada to Dalbir Bindra. The author wishes to thank William Lieberman for assistance in running the experiments.

[2] Now at the University of Auckland, Auckland, New Zealand.

is, recognition should be slower the later the critical digit occurred in the series, and slowest of all when it did not occur in the series.

Somewhat different results might be predicted by Marbe's Law (Osgood, 1953, p. 722), which states that reaction times for free-association responses are inversely related to the frequency of occurrence of these responses. If we assume this to mean that *S* has more rapid access to strong traces than to weak traces, we might expect recognition times to show a recency effect as well as a primacy effect; because early and late items in a serial list are typically recalled and recognized more accurately than items in the middle of the list, it is probable that their traces are stronger, so that access to early and late items should be faster than access to middle items.

Experiment II was designed to provide more direct comparison between recognition time and order of recall. In this experiment three free-recall trials were embedded among the recognition trials. The signal for *S* to recall the digits rather than make a recognition response was not given until after the series of five digits had been presented, so that the trace would have the same properties for recall as for recognition trials. By having *S* make both recall and recognition responses based on essentially identical traces, it was hoped to elucidate more fully the nature of serial processes involved in recall, recognition, and storage.

EXPERIMENT I

Method

Subjects.—The *S*s were 20 volunteers, 10 men and 10 women, from an introductory psychology class. Their ages ranged from 17 to 21 yr.

Apparatus and procedure.—Digits were projected by a multiple-stimulus projector (Grason-Stadler, Model E-4580-1) capable of projecting any of the digits 0 through 9, or a red flash, onto a black screen measuring 2 in. × 1½ in. The duration of digit and red flash presentations, and the duration between successive stimuli, were controlled by an output from a 12-channel tape programming device. Eleven of the 12 channels were connected one-to-one with each of the 11 possible stimuli; the twelfth channel was used to activate two timers. The tape was preprogrammed to select 100 series of eight stimuli, each series consisting of a red flash (which served as a warning signal), five different digits in succession, another red flash, and then a sixth or "critical" digit. In 50 of the 100 series the critical digit was included once among the previous five, in the other 50 series it was not included among the previous five. Each of the 10 digits, 0 through 9, appeared as the critical digit in 10 series; in five of these it had also appeared among the previous five, once in each of the five possible positions. In all but the above respects the digits were randomly selected and the series were randomly ordered.

The *S* was seated directly in front of the projection screen. On an armrest beside him were three buttons arranged at the vertices of an isosceles triangle with the apex towards him. He was instructed to hold the index finger of his preferred hand on the nearest button. When the critical digit appeared he was to decide as quickly as possible whether or not it had appeared among the previous five digits, and to indicate his decision by lifting his finger and placing it on the appropriate one of the other two buttons—a finger movement of about 1 in. For half the *S*s the "yes" button was towards the side of *S*'s preferred hand, while for the remaining *S*s the "no" button was towards this side. When the critical digit appeared two timers were activated. One stopped when *S* lifted his finger, recording "lifting time," the other stopped when *S* pressed one of the decision ("yes," "no") buttons, recording "reaction time."

The *S*s were divided into two equal groups, each containing five men and five women. One group received the stimuli at a "slow" presentation rate in which each digit and red flash was "on" .3 sec. and there was an interval of .3 sec. between offset of one digit or red flash and onset of the next digit or flash in the series. The interval between series was 8 sec. The other group received the stimuli at a "fast" rate, exactly twice the slow rate; that is, "on" and "off" intervals were .15 sec., and there was an interval of 4 sec. between

TABLE 1

PERCENTAGE OF RECOGNITION ERRORS FOR EACH POSITION OF THE CRITICAL DIGIT IN THE PREVIOUS FIVE[a] AND FOR OCCASIONS ON WHICH THE CRITICAL DIGIT DID NOT APPEAR IN THE PREVIOUS FIVE[b], FOR EACH PRESENTATION RATE IN EXP. I AND FOR EXP. II

	Position of Critical Digit in Series					
	1	2	3	4	5	Did not appear
Exp. I						
Slow	12	10	6	6	3	3.2
Fast	13	19	10	2	3	5.8
Exp. II	11	12	6	8	5	5.4

[a] Error of pushing "no" instead of "yes" button.
[b] Error of pushing "yes" instead of "no" button.

series. The *S*s were given three practice series before the experimental series began.

Results

Errors.—Percentages of errors are shown in Table 1. Although errors were infrequent, less than 10% of the total decisions, two trends seem clear. First, there was a recency effect—fewer errors the later the critical digit appeared in the series. Second, there were fewer errors of pushing the "no" instead of the "yes" button than of pushing the "yes" instead of the "no" button.

Lifting and reaction times.—Mean lifting and reaction times were calculated for correct decisions only and are plotted in Fig. 1. For both lifting and reaction times, and both fast and slow presentation rates, there was a clear recency effect. Analyses of variance of correct "yes" response times, as a function of serial position of the critical digit in the series, were carried out separately for lifting and reaction times and for each presentation rate.

In each of the four analyses the linear trends were significant: for the slow presentation rate, $F\ (1,\ 36) = 6.02$, $p < .05$, for lifting times and $F\ (1,\ 36) = 6.94$, $p < .05$, for reaction times; for the fast presentation rate, $F\ (1,\ 36) = 43.11$, $p < .001$, for lifting times and $F\ (1,\ 36) = 54.24$, $p < .001$, for reaction times. It is clear from Fig. 1 that these trends represent recency effects, not primacy effects. The quadratic trend was also significant for lifting times at the slow presentation rate, $F\ (1,\ 36) = 9.18$, $p < .01$. Lifting and reaction times clearly showed very similar trends.

Discussion

There were clear recency effects in both number of errors and response speed; that is, number of errors declined and recognition became faster the later the critical digit occurred in the series. This fails to support the nonassociative model (e.g., Conrad, 1965) which holds that sequentially presented items are held in storage in the order of presentation, and also creates difficulties for the view that short-term storage depends on associative connections between items (Wickelgren, 1966). Rather, it appears that *S* has direct access to each individual item and that speed of access is a

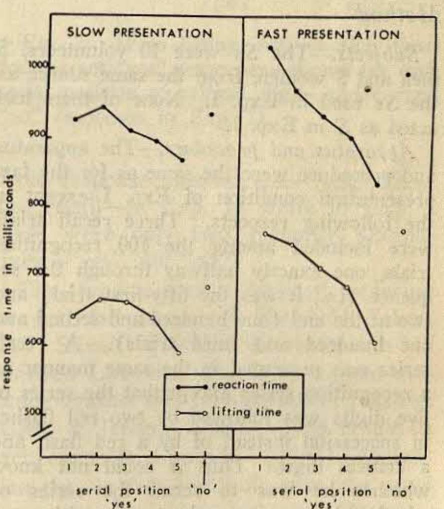


FIG. 1. Mean lifting and reaction times for correct "yes" responses—shown for each serial position of critical digit—and correct "no" responses, for slow and fast presentation rates in Exp. I.

function, as predicted by Marbe's Law, of the strength of the trace of the item and not of its serial position per se. The slight primacy effect which occurred at the slow presentation rate probably resulted because *S* had more time to rehearse at this rate, thus strengthening the traces of earlier items in particular (Corballis, 1966).

Since the results supported Marbe's Law, it seemed of interest to determine whether order of recall would also be a function of the strength of individual traces. Bousfield, Cohen, and Silva (1956), extending Marbe's Law, have presented evidence in support of the theory that in free recall items whose traces are strong are reported before items whose traces are weak. If this theory is true, it follows that if *S*s had been asked to recall the digits in Exp. I they should have most often reported the last digits first; in fact in the fast presentation condition they might have been expected to report the digits in the order exactly opposite to the order in which they were presented. Experiment II was designed to test this possibility.

EXPERIMENT II

Method

Subjects.—The *S*s were 10 volunteers, 5 men and 5 women, from the same source as the *S*s used in Exp. I. None of them had acted as *S* in Exp. I.

Apparatus and procedure.—The apparatus and procedure were the same as for the fast presentation condition of Exp. I except in the following respects. Three recall trials were included among the 100 recognition trials, one exactly halfway through the sequence (i.e., it was the fifty-first trial) and two at the end (one hundred and second and one hundred and third trials). A recall series was presented in the same manner as a recognition series except that the series of five digits was followed by two red flashes in succession instead of by a red flash and a critical digit. Thus *S* could not know whether he was to recall the series or whether he was to make a recognition response until after the presentation of the last digit in the series. Presentation rate was the same for recall series as for recognition series, with "on" and "off" intervals of .15 sec. However, the interval between a recall series and the next series was 8 sec., whereas the interval following a recognition series was 4 sec.

The *S* was instructed to make recognition responses when the critical digit occurred after a red flash, as in Exp. I. In addition, he was told that when he saw a red flash in place of a critical digit he was to attempt to recall the digits, in any order he liked, by speaking them clearly into a microphone which was placed directly in front of him. He was not told how many recall trials he would receive, neither was he given any indication when he would receive them. However, he was given five practice trials, the fourth of which was a recall trial, before the experimental trials began.

The *S*'s recall was fed from the microphone into an amplifier, and from the amplifier into a speaker in another room, where *E* wrote down the responses.

Results

Errors.—Percentages of errors are shown in Table 1 along with those of Exp. I. They show trends very similar to those found in Exp. I.

Lifting and reaction times.—Mean lifting and reaction times for correct decisions are shown in Fig. 2. Again there were clear recency effects, with significant linear trends for both measures, $F\ (1, 36) = 13.57$, $p < .001$, for lifting times and $F\ (1, 36) = 9.57$, $p < .01$, for reaction times. The quadratic component was also significant for lifting times, $F\ (1, 36) = 10.56$, $p < .01$.

Although presentation rate was the same as the fast presentation rate used in Exp. I, the serial position curves in this experiment more closely resemble those of the slow presentation rate condition in Exp. I. Another point of interest is that both lifting and reaction times were somewhat longer than those in Exp. I.

Recall.—Counting number correct regardless of order of recall, the means out of 5 for each of the three recall trials were 4.5, 4.6, and 3.8, respectively. The somewhat lower accuracy for the third recall trial probably occurred because this trial immediately followed the second recall trial and *S*s did not expect to receive two recall trials in succession. Hence they probably did not rehearse as much as in the previous trials.

Order of report was measured by calculating Kendall's tau (Kendall, 1948) between order of report of correct digits and the order of presentation of those digits. Thus a tau score of 1.0 on any recall trial means that *S* reported what digits he could in exactly the order in which they were presented. For the first two recall trials the mean values of tau were .94 and .91, respectively. In each case 8 of the 10 *S*s obtained tau scores of 1.0. For the third recall trial the mean value of tau was only 0.27. This low tau score may be attributed partly to guessing, or attempts to reconstruct information from a weak trace (Conrad, 1965); however, there may also have been some tendency for *S*s to reverse the order over and above the reordering processes inferred by Conrad. One *S*, for example, reported three digits correctly but gave them in the order exactly the reverse of the order in which they had been presented; two *S*s reported the last item first. In the main, however, it is clear that order of recall was predominantly the same as the order of presentation.

Discussion

Recognition times again showed a clear recency effect, supporting Marbe's Law. The slight primacy effect may be interpreted to mean that *S*s rehearsed more in this experiment than did *S*s who received the fast presentation condition in Exp. I, probably because they had been instructed that they would receive some recall trials. The generally longer recognition times in Exp. II may reflect the fact that there was greater response uncertainty in this experiment than in Exp. I (e.g., Dillon, 1966). The *S*s had to choose between three response alternatives—"yes," "no," and recall—in Exp. II while in Exp. I there were only two response alternatives—"yes" and "no."

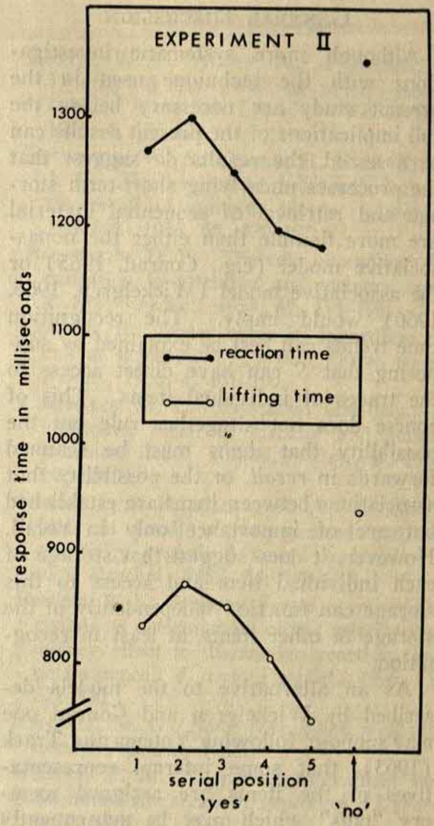


FIG. 2. Mean lifting and reaction times for correct "yes" responses—shown for each serial position of critical digit—and correct "no" responses in Exp. II.

The predominantly forward ordering in recall fails to support Bousfield, Cohen, and Silva's (1956) interpretation of Marbe's Law, and thus adds to the evidence against this interpretation (Battig, Allen, & Jensen, 1965). This result in fact supports Conrad's (1965) contention that in free recall *S*s scan the trace forwards. However, it appears that the forward ordering of items in recall is a function of the recall process itself and not, as Conrad's model implies, a function of storage.

GENERAL DISCUSSION

Although more systematic investigations with the technique used in the present study are necessary before the full implications of the present results can be assessed, the results do suggest that the processes underlying short-term storage and retrieval of sequential material are more flexible than either the nonassociative model (e.g., Conrad, 1965) or the associative model (Wickelgren, 1965, 1966) would imply. The recognition time trends can best be explained by supposing that *S* can have direct access to the traces of individual items. This of course does not altogether rule out the possibility that items must be scanned forwards in *recall,* or the possibility that associations between items are established but are of importance only in recall. However, it does suggest that storage of each individual item and access to this storage can function independently of the storage of other items, at least in recognition.

As an alternative to the models described by Wickelgren and Conrad one may suppose, following Yntema and Trask (1963), that some internal representatives of the items are assigned memory "tags" which may be subsequently scanned by some scanning mechanism. One kind of tag might simply describe whether or not the item has recently been presented. The task of recognizing whether a given item was recently presented would then simply require investigation of the internal representative of the item to determine whether this tag was present, a process which may be similar to that implicated in signal detection tasks.

In recall, Yntema and Trask propose that the scanning mechanism retrieves the items according to how they have been tagged. They assume that the tags representing forward order are more convenient or more reliable than those representing backward order, so that *S* usually reports the items forwards. However, he does not always do so. For example, there appeared to be some tendency to reorder the items in the third recall trial in Exp. II. More compelling examples in which order of report is not the same as the order of presentation are provided by the "channel by channel" ordering strategy in dichotic listening (e.g., Broadbent, 1954) and by Posner's (1964) demonstration that the strategy of reporting the last half of an eight-item list before reporting the first half may be at least as efficient as that of reporting the items forwards from the beginning. The efficiency of any given strategy may depend on the nature of the tags and the descriptive classifications they represent, and on how long the items have been in storage. This latter factor, time in storage, may be particularly important and is worth discussing in detail.

First, following Waugh and Norman (1965), we may distinguish between "primary" and "secondary" storage. Items enter first into primary storage and are likely to be quickly forgotten unless rehearsed, in which case they pass into the more stable secondary storage. The probability of forgetting an item in primary storage is a function of the number of stimulus presentations and responses intervening between presentation of an item and its attempted recall. It follows that the most efficient strategy for attempting to recall all of the items when all are held in primary storage will in general be to scan them forwards, since this strategy will ensure an equal number of intrusions (stimuli and attempted responses) between presentation and recall of each item. Any other ordering strategy will result in a disproportionate number of intrusions between presentation and recall of particular items. Since forward ordering is in general the most efficient ordering strategy when all the items are in primary storage, it is also likely to be the preferred strategy as well as the most practiced one.

When items are retained in secondary storage, it is assumed that their traces are more or less stable and are not subject to significant amounts of disruption from stimulus and response intrusions. When some or all of the items are held in secondary storage, then, there may be no special advantages in forward recall. For example, when the list to be recalled is fairly long or when speed of presentation is fairly slow, the first items may be fairly well rehearsed and therefore retained in secondary storage, while later items are still in primary storage. In such cases the most effective strategy might be to report the later items first in a forward order, thereby minimizing the number of intrusions between presentation and recall of these items, and then subsequently report the earlier, secondary storage items. Posner (1964) has in fact shown that such a strategy becomes more efficient as presentation rate decreases, and Deese and Kaufman (1957) noted ordering strategies of this kind in free recall of word lists that were fairly long (10 words, 32 words, and 50 words). It is possible that Bousfield, Cohen, and Silva's (1956) extension of Marbe's Law to predict serial order in free recall, in which it is argued that order of recall is a function of the strengths of the traces of the items ("strength" tags), applies only to items retained in secondary storage. These authors used long word lists (40 words) presented fairly slowly (one word every 3 sec.).

Finally, it is conceivable that forward recall may not be the most efficient strategy even when the items are stored in primary storage. If we assume that the strength of trace in primary storage may vary, as a function of *S*'s attentiveness, for example, it is possible in certain circumstances that the trace, or identification tag, may be so weak that *S* can be sure of the last item only. In this case if *S* adopts the strategy of reporting the items from the beginning he runs the risk of losing even this last item and thus failing to report any of the items correctly. Hence a better strategy might be to report the last item first, making sure of at least one correct response, and then reconstruct as much as possible of the rest. This could possibly account for the apparent reverse ordering tendency of some *S*s in the final recall trial in Exp. II. These *S*s, not expecting a recall trial immediately following a recall trial, may have failed to establish an effective trace for any but the last digit.

REFERENCES

Battig, W. F., Allen, M., & Jensen, A. Priority of free recall of newly learned items. *J. verbal Learn. verbal Behav.*, 1965, **4**, 175–179.

Bousfield, W. A., Cohen, B. H., & Silva, J. G. The extension of Marbe's Law to the recall of stimulus words. *Amer. J. Psychol.*, 1956, **69**, 429–433.

Broadbent, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, **47**, 191–196.

Conrad, R. Order error in immediate recall of sequences. *J. verbal Learn. verbal Behav.*, 1965, **4**, 161–169.

Corballis, M. C. Rehearsal and decay in immediate recall of visually and aurally presented items. *Canad. J. Psychol.*, 1966, **20**, 43–51.

Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

Dillon, P. J. Stimulus versus response decisions as determinants of the relative frequency effect in disjunctive reaction time performance. *J. exp. Psychol.*, 1966, **71**, 321–330.

Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.

Milner, P. M. A neural mechanism for the immediate recall of sequences. *Kybernetik*, 1961, **1**, 76–81.

Norman, D. A., & Wickelgren, W. A. Short-term recognition memory for single digits and pairs of digits. *J. exp. Psychol.*, 1965, **70**, 479–489.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford, 1953.

Posner, M. I. Rate of presentation and order of recall in immediate memory. *Brit. J. Psychol.*, 1964, **55**, 303–306.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychol. Rev.*, 1965, **72**, 89–104.

Wickelgren, W. A. Short-term memory for repeated and non-repeated items. *Quart. J. exp. Psychol.*, 1965, **17**, 14–25.

Wickelgren, W. A. Phonemic similarity and interference in short-term memory for single letters. *J. exp. Psychol.*, 1966, **71**, 396–404.

Yntema, D. B., & Trask, F. P. Recall as a search process. *J. verbal Learn. verbal Behav.*, 1963, **2**, 65–74.

(Received May 23, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 106–113

CONNOTATIVE MEANING IN VERBAL LEARNING[1]

SUSAN PATTON HARBIN AND JOHN H. WRIGHT
Wake Forest College *Virginia Polytechnic Institute*

In Experiment I CVC trigrams from a narrow range of medium association values were scaled on the activity, potency, and evaluative dimensions of connotative meaning. In 3 subsequent experiments Ss learned 8-item PA or serial lists of CVCs differing in connotative meaning. CVC trigrams high in connotative activity were more difficult to learn than CVCs potent or neutral in connotative meaning but potent CVCs were not more difficult than neutral CVCs. No differences were found in the retention of connotatively meaningful and neutral CVCs. These results were interpreted in terms of previously reported differential correlations between activity and potency ratings and associative uncertainty.

Johnson, Thomson, and Frincke (1960) have reported a positive relationship between the association values of CVC trigrams and the evaluative ratings of these trigrams on the semantic differential. Syllables of high association value were generally judged "good," whereas syllables of low association value were generally judged "bad," a finding which suggested that connotative meaning may have considerable relevance for verbal learning. Employing the items rated by Johnson et al. (1960), Keppel (1963) investigated the evaluative dimension of connotative meaning in verbal learning and found that "good" syllables were learned significantly faster than "bad" syllables. However, connotative evaluation was largely confounded with pronunciability in this study—"bad" syllables were generally more difficult to pronounce than were "good" syllables. Thus, the role of connotative meaning in verbal learning remains obscure.

The purpose of the experiments to be reported was to investigate connotative meaning in paired-associate (PA) and serial verbal learning. The learning materials in these experiments were CVC trigrams selected from an extremely narrow range of association values (48% through 52% according to Archer, 1960) in order to hold associative meaningfulness constant while permitting the connotative meaning of PA and serial lists to vary. The selection of learning materials from this range of association values also served to control pronunciability. Lindley (1963) has shown that the serial learning of CVC trigrams taken from the middle range of association values is not related to pronunciability.

EXPERIMENT I

The purpose of this experiment was to obtain semantic-differential ratings of CVC trigrams taken from the middle range (48% through 52%) of association values reported by Archer (1960).

Method

The *S*s, 40 males and 40 females enrolled in introductory psychology classes at Wake Forest College, rated the 106 trigrams from the specified range of association value on a form of Osgood's semantic differential. In order to reduce the length of the rating task half of the syllables were assigned to each of two lists. Each *S* rated only the syllables in one list, and the number of males and females rating each list was equated. The semantic-differential booklet contained a cover sheet with instructions and a separate rating sheet for each of the 53 syllables in the list. Each rating sheet consisted of the same 12 scales, and each scale was divided into seven equal sections with bipolar adjectives placed at the extremes. The scale adjectives were taken from Osgood, Suci, and Tannenbaum (1957), each set of bipolar adjectives having been shown by factor analysis to be maximally loaded on one of three dimensions of connotative meaning—evaluation (E), potency (P), and activity (A). The scales employed to assess each dimension were as follows: The E scales were good-bad, nice-awful, foul-fragrant, clean-dirty, unpleasant-pleasant, and sacred-profane; the P scales were large-small, strong-weak, and heavy-light; the A scales were dull-sharp, active-passive, and slow-fast. The *S*s were instructed that previous research had indicated that people have certain "feelings" about nonsense syllables and that their task would be to rate their impressions or feelings about each syllable. The *S*s were urged not to be too "intellectual" in their ratings but to respond with their immediate reactions.

---

[1] This study is based upon a thesis submitted to the graduate faculty of Wake Forest College in partial fulfillment of the requirements for the MA degree by Susan Patton Harbin under the direction of John H. Wright. The assistance of John E. Williams in the collection and analysis of the data of Exp. I is gratefully acknowledged.

TABLE 1

*M*s AND *SD*s OF SEMANTIC-DIFFERENTIAL RATINGS OF CVC TRIGRAMS MEANINGFUL ON ONE OR MORE DIMENSIONS OF CONNOTATIVE MEANING

Syllable	Dimensions of Connotative Meaning					
	Activity		Potency		Evaluation	
	M	*SD*	*M*	*SD*	*M*	*SD*
BYG	3.24	1.44	5.66	1.31	4.19	1.11
DUH	1.65	0.80	4.92	1.56	4.64	0.70
FOT	4.10	1.44	4.34	1.10	5.33	0.92
GEY	5.02	1.23	3.91	1.30	2.92	1.00
GUK	2.85	1.34	5.29	1.29	6.02	0.72
HAX	5.22	1.45	5.39	1.01	4.46	0.91
HEF	4.17	1.40	5.47	1.38	3.96	0.90
JEY	5.02	1.42	3.78	1.22	3.65	1.10
JIV	5.79	1.12	3.74	1.06	3.51	0.79
KAC	5.15	1.33	4.63	1.29	4.14	1.40
KES	4.55	1.48	3.74	1.24	2.61	0.88
KYL	5.14	1.20	4.59	1.30	4.86	1.44
KYN	4.72	1.13	4.10	1.32	2.98	0.96
KYX	5.64	1.41	4.08	1.30	3.46	1.07
LAH	4.16	1.78	3.76	1.18	2.84	1.02
MUW	2.95	1.28	4.70	1.59	4.28	1.00
MUZ	2.75	1.32	4.49	1.16	4.26	1.25
NOH	3.77	1.63	5.04	1.12	3.86	1.25
NYK	5.44	1.32	3.77	1.46	4.14	1.24
QIK	5.09	1.68	3.92	1.31	3.52	0.91
ROQ	3.76	1.48	5.29	0.96	4.45	0.99
TAQ	5.54	3.75	3.96	2.44	3.93	1.24
VEP	5.42	1.53	3.50	1.12	3.42	0.82
VIZ	5.72	1.27	3.74	1.04	3.58	1.12
VOD	3.94	1.90	5.45	1.74	4.84	1.36
VOS	4.25	2.84	5.19	4.19	4.01	1.20
VUL	3.46	1.62	5.23	1.35	5.28	1.19
WOH	3.45	1.82	5.07	1.13	3.99	1.04
WYF	5.04	1.37	3.54	1.45	3.06	1.32
ZAR	4.66	1.43	5.26	1.20	4.25	1.08
ZEK	5.63	3.77	5.11	4.26	3.95	1.12

Results

Mean ratings for each syllable on each of the three dimensions of connotative meaning were computed separately for males and females, but since no sex differences were evident the data were combined. The mean ratings for each syllable on the three A scales were averaged to obtain the A value for that syllable, and the E and P values were obtained in a similar fashion by averaging the mean ratings for each syllable on the six E scales and the three P scales. Syllables with mean ratings between 3.00 and 5.00 on all three dimensions were classified as connotatively neutral, whereas syllables with mean ratings less than 3.00 or greater than 5.00 on one or more of the dimensions were classified as connotatively meaningful. Thirty-one of the 106 trigrams rated on the semantic differential were found by this criterion to be meaningful on at least one of the three dimensions of connotative meaning (Table 1). Twenty-six of these 31 CVCs were rated meaningful on a single dimension (11 high A, 3 low A; 8 high P; 1 high E, 3 low E), four of the 31 were meaningful on two dimensions, and only one trigram was meaningful on all three dimensions. The reliabilities of the A, E, and P values presented in Table 1 were estimated by randomly dividing *S*s rating a particular syllable into two groups containing equal numbers of males and females. The correlations between the mean ratings of the different subgroups for the 31 syllables in Table 1 were computed separately for the A, E, and P dimensions and were .85, .89, and .48 for these dimensions, respectively, all of which were highly significant, t (29) ≥ 2.982, $p < .01$.

EXPERIMENT II

The learning of paired associates differing in connotative meaning was investigated in this experiment.

Method

Eighteen males and 12 females from introductory psychology classes at Wake Forest College learned a single eight-pair PA list of minimal interitem similarity (repeated letters) constructed from the CVC trigrams scaled for connotative meaning in Exp. I. The list contained four pairs of syllables rated neutral on all three dimensions of connotative meaning and four pairs of syllables rated high in meaning on one of the dimensions but neutral on the two remaining dimensions. Three of the four connotatively meaningful pairs were high in potency and the remaining meaningful pair was high in activity.

The recall method of PA learning, wherein all S-R pairs are presented for learning prior to S-term presentation for testing, was employed. During testing *S* spelled aloud the three letters of the PA response terms. Learning was carried to a criterion of one errorless trial but was discontinued if this criterion was not attained within a 30-min. experimental session. A 2-sec. rate of presentation was employed with 4 sec. lapsing between the learning and test portions of a trial. Intertrial intervals were approximately 10 sec. The order of presentation of S-R pairs and S-terms was different within a single trial and also varied unsystematically from trial to trial.

TABLE 2

*M*s AND *SD*s OF ERRORS PER CONNOTATIVELY MEANINGFUL AND NEUTRAL PAIRED ASSOCIATE

Performance Measures	Connotative Meaning	
	Meaningful	Neutral
Before Errors		
M	4.13	3.94
SD	3.12	3.11
After Errors		
M	1.40	0.89
SD	1.84	1.33
Total Errors		
M	5.53	4.83
SD	3.38	3.35

Results

Presented in Table 2 are mean errors occurring before and after the first correct response to each stimulus and mean total errors for the connotatively meaningful and neutral pairs within the list. Analysis of variance revealed that the meaningful pairs were more difficult to learn than were the neutral pairs, for total errors, F (1, 29) $= 6.70$, $p < .025$. This effect resulted primarily from the greater mean number of after errors per meaningful than per neutral pair, F (1, 29) $=$ 6.21, $p < .025$. The difference in mean number of before errors per pair, although in the same direction, failed to attain statistical significance ($F <$ 1.00).

EXPERIMENT III

This experiment attempted to extend the results of Exp. II by manipulating single dimensions of connotative meaning in both the stimulus and response positions within PA lists. An attempt was also made to determine if there are differences in the retention of paired associates varying along the separate dimensions of this variable.

Method

Subjects.—Thirty-six male and 36 female students enrolled in introductory psychology classes at Wake Forest College served as *S*s in order to satisfy a course requirement.

Materials.—Six eight-pair PA lists were constructed from the CVC trigrams scaled in Exp. I. Three of the PA lists contained the same eight highly potent syllables (i.e., syllables with a mean rating greater than 5.00 on the P dimension) and the same eight neutral syllables (i.e., syllables with mean ratings between 3.00 and 5.00 on all three meaning dimensions). Six of the eight potent syllables were neutral on the A and E dimensions. One of the two remaining potent CVCs rated high on the A dimension and one rated high on the E dimension. The three remaining PA lists contained the same eight highly active syllables (i.e., syllables with mean ratings greater than 5.00 in activity) and a second set of eight neutral syllables. With the exception of one syllable low in evaluation the active syllables were neutral on the P and E dimensions. Due to the limited number of syllables having comparable degrees of evaluative meaning, it was not possible to construct in a similar fashion highly evaluative lists.

Four different types of pairings were present within each active and potent list. Two pairs in each list contained connotatively meaningful syllables in both the stimulus and response positions (M-M pairs), while a second two pairs contained connotatively neutral syllables in both positions (N-N pairs). Two of the remaining four pairs consisted of connotatively meaningful stimuli and neutral responses (M-N pairs), and two consisted of neutral stimuli and connotatively meaningful responses (N-M pairs).

In order to minimize interitem repetition of consonants and vowels it was necessary to employ the one active and two potent syllables connotatively meaningful on one additional dimension. In addition, the construction of three different active and three different potent lists served to equate intralist stimulus similarity, intralist response similarity, and intralist S-R similarity for the four kinds of pairings within both active and potent lists and also to equalize any facilitative or detrimental effects which might result from specific S-R pairings.

Experimental design.—Eight *S*s learned each of the nine possible combinations of one active and one potent list. Four of the *S*s in each active-potent list-combination condition learned the active followed by the potent list and the remaining four *S*s learned the potent followed by the active list. An equal number of males and females were assigned to these 18 conditions on a randomized basis in order of appearance for the experiment.

TABLE 3

*M*s AND *SD*s OF TOTAL ERRORS FOR THE FOUR PAIR CATEGORIES WITHIN ACTIVE AND POTENT LISTS

Lists	Pair Category				*M*
	M-M	M-N	N-M	N-N	
Active					
M	15.04	12.33	12.39	9.25	12.25
SD	4.80	4.67	4.25	3.50	
Potent					
M	13.88	13.90	13.96	12.99	13.68
SD	4.84	4.84	4.44	4.42	
M	14.46	13.12	13.17	11.12	

Procedure.—The procedure for PA learning was identical to that of Exp. II with *S* spelling aloud the three letters of the PA response terms during the test portion of a trial. Learning was carried to a criterion of one errorless trial but was discontinued following Trial 20 if this criterion had not been attained. One week following List-1 learning *S*s returned for a second experimental session which for half the *S*s in each condition consisted of two measures of List-1 retention followed immediately by List-2 learning. The remaining half of the *S*s received only List-2 learning in the second experimental session in order to evaluate the possible effects of the List-1 retention tests on List-2 learning. However, the equivalence of List-2 learning for these groups indicated the absence of an effect of the List-1 retention tests upon List-2 learning and subsequent analyses are based upon the combined performance of these groups. The first of the retention tests, a free-recall measure, required *S* to write down all stimulus and response terms from List 1. The second test of retention was a recognition matrix consisting of 16 rows of eight items each. One CVC in each row was a stimulus or response term from List 1 and the remaining seven items in the row were CVCs highly similar in letter content. The position of rows containing a stimulus or response item, as well as the position of the single List-1 item within each row, was randomized in the matrix. The *S*s were instructed to identify the 16 items which had previously appeared in List 1. Following the List-1 retention tests *S*s learned the second PA list, and one week following List-2 learning free-recall and recognition tests of List-2 retention were given to all *S*s.

Results

Paired-associate learning.—Because of the failure of 18 *S*s to reach the one-errorless-trial learning criterion by Trial 20 the after-error measure was generally insensitive to the effects of the variables manipulated in Exp. III, and for this reason only total errors are considered. The mean number of total errors for each of the four pair categories within the active and potent lists is presented in Table 3 and analysis of variance of this measure is summarized in Table 4. Lists containing potent syllables were significantly more difficult to learn than were lists containing active syllables ($p < .005$). Inspection of Table 3 reveals that the pairings effect was greater in the active than in the potent lists, thus resulting in a significant Pairings × A-P interaction ($p < .005$). Additional analysis revealed no differences among the M-M, M-N, and N-M pairs within the potent lists ($F < 1.00$), but there was a tendency for these pairings to produce more errors than the N-N pairings within these lists, $F\ (1, 953) = 3.23$, $p < .10$. For the active lists, on the other hand, M-M pairs were more difficult than the M-N and N-M pairs and these pairs in turn were more difficult than N-N pairs, $F\ (1, 953) \geq 24.04$, $p < .001$.

TABLE 4

ANALYSIS OF VARIANCE OF TOTAL ERRORS IN THE FOUR PAIR CATEGORIES WITHIN ACTIVE AND POTENT LISTS

Source	*df*	*MS*	*F*
Between *S*s	71		
A-P List Combinations	8	88.90	< 1.00
A-P—P-A List Sequences	1	159.76	< 1.00
Combinations × Sequences	8	116.77	< 1.00
Error between	54	174.03	
Within *S*s	1080		
Lists 1–2	1	540.38	38.18***
A-P Lists	1	146.64	10.36**
Pair Categories (PC)	3	140.00	9.77***
Stimulus M-N	1	250.57	14.48***
Response M-N	1	194.21	13.55***
Stimulus M-N × Response M-N	1	9.21	< 1.00
PC × A-P Lists	3	72.60	5.07**
Stimulus M-N × A-P Lists	1	108.16	7.55*
Response M-N × A-P Lists	1	108.15	7.55*
Stimulus M-N × Response M-N × A-P Lists	1	1.45	< 1.00
PC × Lists 1–2	3	10.22	< 1.00
PC × A-P Combinations	24	22.55	1.55
PC × A-P—P-A Sequences	3	14.95	1.04
PC × Combinations × Sequences	24	14.05	< 1.00
Between Cells Residual	65	14.58	1.02
Error within	953	14.33	

* $p < .01$.
** $p < .005$.
*** $p < .001$.

Connotatively meaningful syllables were more difficult to learn than connotatively neutral syllables both as PA stimuli and as PA responses ($p < .001$), and furthermore, these effects were of identical magnitude (for the Stimulus M-N × Response M-N interaction $F < 1.00$). In addition, the effects of connotative meaning in the stimulus and response positions were greater in the active than in the potent lists (for the Stimulus M-N × A-P and Response M-N × A-P interactions $p < .01$). The effects of connotative meaning in the stimulus position of the active lists were equivalent to the effects of this variable in the response position of these lists, and the effects of this variable in the stimulus and response positions of the potent lists, although of smaller magnitude, were also equivalent (for the Stimulus M-N × Response M-N × A-P interaction $F < 1.00$).

Paired-associate retention.—Analysis of variance of mean number of trigrams free recalled and recognized correctly in retention revealed no significant differences in the retention of connotatively meaningful and neutral trigrams within the active and potent lists ($p > .10$). The only significant effects in retention were the better free-recall and recognition of PA responses than of PA stimuli ($p < .001$).

EXPERIMENT IV

The purpose of this experiment was to investigate the learning of serial lists

TABLE 5

*M*s AND *SD*s OF PERFORMANCE IN CONNOTATIVELY ACTIVE, POTENT, AND NEUTRAL SERIAL LISTS

Performance Measures	Connotative Meaning		
	Active	Potent	Neutral
Trials to Criterion			
M	12.86	11.14	10.14
SD	5.74	4.29	3.52
Before Errors			
M	34.11	29.47	26.81
SD	10.89	10.48	8.56
After Errors			
M	14.92	10.92	10.11
SD	11.73	8.11	8.22
Total Errors			
M	49.03	40.39	36.92
SD	21.93	17.37	18.97

of CVC trigrams differing in connotative meaning.

Method

Eighteen male and 18 female psychology students at Wake Forest College learned by the method of serial anticipation three successive eight-item lists. Items were presented at a 2-sec. rate and intertrial intervals were 6 sec. Each list was learned to a criterion of one errorless trial and a 10-min. rest period was given between successive lists. During learning *S*s spelled aloud the three letters of the trigrams comprising the lists. An asterisk served as a cue for the first item of the list.

A total of nine different serial lists was employed. Three of the lists were constructed from the same eight active syllables, three from the same eight potent syllables, and the remaining three lists were constructed from the same eight neutral syllables. The active syllables were neutral on the P and E dimensions, the potent syllables were neutral on the A and E dimensions, and the neutral syllables were neutral on all three dimensions of connotative meaning. All lists were equivalent in number of repeated consonants and vowels. The construction of three different A, P, and N lists served to distribute syllables and repeated letters uniformly over the eight serial positions. The six sequences of one active, one potent, and one neutral list, along with the counterbalancing of the three different orders of A, P, and N syllables over the three lists, yielded a total of 18 experimental conditions to each of which two *S*s (one male and one female) were assigned on a random basis in order of appearance at the laboratory.

Results

Presented in Table 5 are mean trials to criterion, mean errors occurring before and after the first correct anticipation of each item in the list, and mean total errors for serial lists of active, potent, and neutral trigrams. Analysis of variance revealed the differences among lists in before errors and total errors to be statistically significant, $F\ (2, 36) \geq 4.09,\ p < .05$, but list differences in after errors failed to attain significance, $F\ (2, 36) = 1.65,\ p > .20$. In addition, differences among lists in trials to criterion approached statistical significance, $F\ (2, 36) = 3.19,\ p < .10$. Further analysis revealed these effects to result primarily from the inferiority of performance in the active lists, for the A—P + N comparisons in before errors, total errors, and trials to criterion $F\ (1, 36) \geq 5.53,\ p < .05$, and in after errors $F\ (1, 36) = 3.23,\ p < .10$. None of the differences between potent and neutral lists were significant $(F < 1.00)$.

DISCUSSION

From the results of the present study it is clear that the learning of PA and serial lists of CVC trigrams of medium association value is significantly impaired if these CVCs are connotatively active but is not appreciably impaired if these items are connotatively potent. Since rate of verbal learning is directly rather than inversely related to meaningfulness defined from association values and other measures, the present findings render invalid speculation that connotative meaning as defined by Osgood and meaningfulness defined in terms of traditional measures represent comparable manipulations of kind of learning material. Furthermore, the statistical effects of connotative meaning in the stimulus and response positions of PA lists were equivalent, whereas meaningfulness generally exerts a much more pronounced effect in the response than in the stimulus position (Underwood & Schultz, 1960).

The present findings of an inverse relationship between rate of learning and connotative activity, but not connotative potency, are consistent with a recent analysis of connotative meaning by Howe (1965). Howe reported that words which are more polarized on the semantic differential, i.e., further from the origin or neutral point in their rated connotative meaning, ". . . will tend to elicit higher associative uncertainty, a larger number of different associations, a weaker primary, and a larger number of unique responses than will a word lying closer to the arbitrary Origin of the factor space . . . [p. 507]." Howe's analysis further implies that associative uncertainty is a measure of the degree to which the associations to a verbal item fluctuate over time in their relative positions within the associative hierarchy. He also found that activity and evaluative semantic-differential ratings were highly correlated with these measures of associative uncertainty, whereas potency ratings failed to relate to these measures. These findings suggest that highly active syllables should be more difficult to learn than neutral syllables because of greater associative interference resulting from the greater moment-to-moment hierarchical fluctuation of associations evoked by the active syllables than by the less polarized neutral syllables. However, no differences in the learning of potent and neutral syllables should occur since the potency dimension does not correlate with measures of associative uncertainty. Thus, Howe's analysis seems to explain why active syllables were more difficult than potent or neutral syllables and also why the learning of potent and neutral syllables failed to differ. The distinction between associative meaningfulness and associative uncertainty can now be clarified. Whereas associative meaningfulness is an index of the number of associations to a verbal item and is positively correlated with rate of learning, all of Howe's measures suggest that associative uncertainty reflects the temporal instability of the associations to a verbal item, and for this reason associative uncertainty should be negatively correlated with rate of learning.

Since active words were found by Howe (1965) to possess considerably greater associative uncertainty than potent words, the finding in Exp. III of greater difficulty of potent than active lists at first seems surprising. However, when the difficulty of learning pairs containing active or potent trigrams is evaluated relative to the control pairs in these lists (viz., the N-N pairs), it is apparent that, consistent with Howe's findings, active trigrams were more difficult to learn than were potent trigrams. Inspection of Table 3 reveals that pairs containing active trigrams were more difficult to learn than were the N-N pairs within the active lists, whereas pairs containing potent trigrams were not appreciably more difficult than the N-N pairs within the potent lists. The faster learning of the active than of the potent lists may well have resulted from a pair-selection strategy commonly observed when PA lists differ markedly in heterogeneity of pair difficulty (Postman, 1963). The *S*s learning a PA list of heterogeneous pair difficulty frequently concentrate upon easy pairs early in PA learning and do not attempt to learn the more difficult pairs until much later, whereas *S*s learning a PA list of more homogeneous pair difficulty tend to distribute their learning efforts more uniformly over all pairs of the list during all stages of practice. Since active pairs were more difficult than their respective N-N control pairs whereas potent pairs were not, it seems reasonable to presume that when learning active PA lists *S*s concentrated to a greater extent upon the control pairs early in learning than was the case when learning the potent lists. This contention was supported by the greater difference in mean number of silent errors prior to the first correct response between the active-list N-N pairs and pairs containing active trigrams than between the potent-list N-N pairs and pairs containing potent trigrams. Furthermore, there were fewer silent before errors for the N-N pairs in the active lists than for the N-N pairs in the potent lists, but more silent before errors for the pairs containing active trigrams than for the pairs containing potent trigrams ($p < .05$). Concentration upon easy pairs early in PA learning may in effect reduce the length of the list and, as a consequence, facilitate the overall rate of PA learning. Thus, differences in pair-selection strategy may well account for the finding that potent lists were more difficult to learn than active lists although active trigrams were more difficult to learn than were potent trigrams.

The finding of no differences in the retention of active, potent, and neutral trigrams is not surprising inasmuch as most studies have shown that kind of learning material has no effect upon verbal retention. The better retention of PA responses than of PA stimuli is in agreement with previous findings and provides further evidence that PA response-learning requirements are substantially greater than PA stimulus-learning requirements.

REFERENCES

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.,* 1960, **74** (10, Whole No. 497).

Howe, E. S. Uncertainty and other associative correlates of Osgood's D_4. *J. verbal Learn. verbal Behav.,* 1965, **4**, 498–509.

Johnson, R. C., Thomson, C. W., & Frincke, G. Word values, word frequency, and visual duration thresholds. *Psychol. Rev.,* 1960, **67**, 332–342.

Keppel, G. Word value and verbal learning. *J. verbal Learn. verbal Behav.,* 1963, **1**, 353–356.

Lindley, R. H. Association value, familiarity, and pronunciability ratings as predictors of serial verbal learning. *J. exp. Psychol.,* 1963, **65**, 347–351.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

Postman, L. One-trial learning. In C. N. Cofer (Ed.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963. Pp. 295–321.

Underwood, B. J., & Schultz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received May 26, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 114–123

TEMPORAL RELATIONSHIP BETWEEN RESPONSE AND PUNISHMENT[1]

DAVID S. CAMP,[2] GEORGE A. RAYMOND,[3] AND RUSSELL M. CHURCH
Brown University

In Experiment I (N = 88) the magnitude of response suppression was greater for contingent punishment than for punishment delayed by 30 sec., and it was positively related to the intensity of the punishment (.1, .2, .3, .5, and 2.0 ma.). In Experiment II (N = 45) a systematic delay-of-punishment gradient (0.0, 2.0, and 30.0 sec.) was described in a free-responding situation. In Experiment III (N = 30) no difference in magnitude of response suppression or in the frequency distribution of response-shock intervals was detected between a group with a 30-sec. delay-of-punishment and a group with noncontingent shock. In Experiment IV (N = 84) a delay-of-punishment gradient (0.0, 2.0, 7.5, 30.0 sec., and noncontingent shock) was described in a discriminative situation. All of these results are consistent with the empirical generalization that the magnitude of suppression of an appetitive response is greater if the punishment is contingent upon the response than if it is not.

Punishment of an appetitive response results in response suppression, but a similar response suppression may also be produced by the mere occurrence of aversive stimuli. To obtain an adequate understanding of the mechanisms underlying response suppression it is necessary to determine the extent to which the effects of punishment are based on the temporal relationship between response and punishment (Church, 1963).

We considered three methods of approach to the problem of determining the amount of suppression due to the response dependency of the punishing stimuli. The first was the *noncontingent* method, which has been used in studies of punishment by Azrin (1956), Estes (1944), and Hunt and Brady (1955). With this method the behavior of experimental *S*s punished immediately following some response is compared with that of control *S*s that receive an equivalent aversive stimulus independently of any particular response. A difficulty is that the control *S*s may differ from the experimental *S*s with respect to (*a*) the number of aversive stimuli received and (*b*) the relationship between the number of responses and the number of aversive stimuli received, as well as (*c*) the temporal relationship between response and shock.

The second method was the *yoked control* method, which has been used by Hoffman and Fleshler (1965). In this procedure *S*s are matched according to some relevant criteria before the punishment treatment is begun. In a given matched pair of *S*s, the experimental member receives punishment immediately following some particular response; the control member simultaneously receives the same aversive stimulus regardless of its behavior. This method, unlike the noncontingent method, equates experimental and control *S*s with respect to the number of aversive stimuli received and their distribution through the session. Unfortunately, a difference in degree of suppression may occur simply as a result of individual differences in resistance to shock, or moment-to-moment variability of an individual *S* (Church, 1964).

---

[1] Experiment I formed a portion of the dissertation of David S. Camp in partial fulfillment of the requirements for the doctoral degree at Brown University. Experiment II formed a portion of the dissertation of George A. Raymond in partial fulfillment of the requirements for the Master of Science degree at Brown University. These experiments were supported by a United States Public Health Service Predoctoral Fellowship to David S. Camp and a United States Public Health Service Research Grant (MH 08123) from the National Institute of Mental Health to Russell M. Church.

[2] Now at the University of Rhode Island.

[3] Now at the University of British Columbia.

The third approach was the method of *delay*. A systematic delay procedure has been used in the punishment extinction of avoidance responses (Kamin, 1959) and in discrimination learning (Bevan & Dukes, 1955; Warden & Diamond, 1931). The method involves the use of two or more independent groups of *S*s, each of which receives a different delay of punishment following the response. The assumption is that, as the temporal interval between response and punishment increases, the association between response and punishment decreases while the emotional consequences of the aversive stimulus remain constant. A limiting case involves a single extremely long delay.

The assumption is that the long-delay group is equated with the contingent group with respect to the relationship between the number of responses and the number of aversive stimuli received, but that *S*s in the long-delay group are not likely to form an association between the response that produces the punishment and the punishment.

The present study, involving the noncontingent method and the method of delay, represents an attempt to gain further information concerning the role of the temporal relationship between response and subsequent punishment. The results of all the experiments support the hypothesis that an aversive stimulus occurring immediately after a response (punishment) produces greater suppression of the response than does an aversive stimulus which is temporally removed from the response.

GENERAL METHOD

Subjects.—The *S*s were male albino rats of the Sprague-Dawley strain that arrived from the Charles River Breeding Laboratories at 60 days of age. They were about 90 days old at the start of the experiments, and their weights ranged between 230 and 255 gm.

Apparatus.—All experiments were run in lever boxes 9⅛ in. long, 8 in. wide, and 8⅝ in. high. The ends of the boxes were constructed of aluminum, and the top and sides were made of clear plastic. The floors were constructed of 16 stainless-steel bars $\frac{5}{32}$ in. in diameter. The levers were made of stainless steel and were ½ in. thick and 2 in. wide. Each box was enclosed in a picnic-type ice chest which also contained a pellet dispenser, a blower, and a 7½-w. incandescent lamp mounted directly above the lever box. In Exp. IV a click with a frequency of about 10 clicks/sec and about 78-db. SPL was presented through a speaker located outside of the lever box at the end near the lever. The grid floors were electrified by means of a constant current stimulator which was delivered through scramblers (Applegate Model 250 and Lehigh Valley Electronics Model 1311 in Exp. I and Grason-Stadler Model E1064GS in Exp. II, III, and IV). In Exp. I the intensity of the shock was varied and the duration was 2 sec.; in Exp. II, III, and IV all shocks were .25 ma. for 1-sec. duration. The stimulators and scramblers were located next to the lever boxes; the remaining control and recording apparatus was located in a separate room.

Procedure.—The *S*s were housed in individual cages with water always accessible, and they were handled daily. For the first 4–7 days *S*s were given ad lib food; thereafter, they were maintained on a diet of about 12 gm. of dry Purina chow mixed with about 25 cc. of water. On the first experimental day all *S*s were trained to approach the food cup following the activation of the pellet dispenser by automatically delivering one 45-mg. Noyes food pellet each minute until 30 pellets had been delivered. During this time the lever was present in the box but it could not activate the pellet dispenser. On the second experimental day *S*s were reinforced for each lever-response and were removed after making 30 responses. Thereafter all sessions were 30 min. in length and each *S* was given a daily session at approximately the same time each day. During the following eight sessions of Exp. I and five sessions of Exp. II and III *S*s were reinforced on a 1-min. variable-interval schedule. Punishment procedures were then introduced.

Measure.—The number of responses for each *S* for each of the 10 sessions of punishment training was transformed into a *suppression ratio* by dividing the number of responses on a given session of punishment training by the sum of the number of responses on the given session of punishment training and the number of responses on the last day of reinforcement training.

EXPERIMENT I

To compare the magnitude of suppression produced by punishment that immediately follows a response with that produced by punishment temporally removed from a response, Exp. I used the method of *long delay*. A discrete trials procedure introduces an additional complexity into the problem of the influence of the temporal relationship between response and punishment on response suppression. Since the trial must be begun by some external cue that can serve as the basis for the formation of a conditioned emotional response, observed suppression could be a result of the interval between *cue* and shock rather than the interval between *response* and shock. To eliminate this confounding variable, Exp. I used a free-responding situation.

If differences in the amount of suppression exist between immediate and long-delayed punishment conditions, the detection of these differences must depend on the intensity of the shock. Extremely mild intensities should not substantially alter the response rate under either immediate or delayed conditions, while extremely intense punishments may result in a suppression of all instrumental activity under either of these conditions. Therefore, the delay parameter should be most prominent at moderate punishment intensities.

Although the differential effect of intensity on contingent and long-delayed punishment has not been explored, the effect of intensity on response rate has been investigated extensively (e.g., Azrin, 1960; Karsh, 1962). In all of these experiments the measure of response speed was inversely related to the intensity of the punishment. In the present experiment five punishment intensities were used for different groups of *S*s, ranging from a mild .1 ma. to an intense 2.0 ma. The *S*s in the contingent group were punished immediately following a response that met a particular criterion, and *S*s in the long-delay groups were punished 30 sec. following the criterion response.

Experimental design.—The 88 *S*s used in this experiment were divided into 11 groups of 8 *S*s. One of these was a control group which received no punishment. Of the remaining 10 experimental groups half received immediate punishment (contingent), and half received punishment 30 sec. following a particular response (long-delay). Each of the five contingent groups and each of the five long-delay groups received one of five punishment intensities (.1, .2, .3, .5, or 2.0 ma.).

Procedure.—Following eight sessions of variable-interval reinforcement training *S*s received 10 sessions of punishment training. During punishment training the reinforcement schedule remained in effect (VI-1) and, in addition, punishments of 2.0-sec. duration were delivered according to a fixed-ratio schedule. The ratio for each *S* was determined by dividing its number of responses in the final session of reinforcement training by 30, so that *S* would receive an average of one punishment per minute if its response rate remained unchanged. (Determination of the ratio in this manner assured that equal proportional increases or decreases in the response rates of different *S*s resulted in an equal number of punishments.) Half of the experimental *S*s received a punishment immediately after completing the ratio, and half received a punishment 30 sec. after completing the ratio.

Results.—Figures 1 and 2 show the suppression ratios for all experimental and control groups for the 10 punishment sessions. (The ordinate scale of Fig. 2 is double that of Fig. 1 in order to make the functions more clearly distinguishable.) Inspection of Fig. 1 and 2 indicates that the greater the intensity of the punishment the greater the suppression of response. The mean suppression ratio for the 10 days of punishment training was inversely related to punishment intensity for both the contingent and the long-delay groups, $H(4) = 28.8$, $p < .001$ and $H(4) = 27.5$, $p < .001$.[4] Individual intensity comparisons for the Contingent and Long-Delay groups were performed using the Mann-Whitney *U* test (two-tailed test with $\alpha = .05$). In both cases the 2.0-ma. *S*s and the .5-ma. *S*s differed from each other and from all other intensity groups. The .3-ma. *S*s at long delay were more suppressed than the .2-ma. *S*s at long delay. The control group was relatively less suppressed than any group except the .1-ma. and .3-ma. contingent groups and the .1-ma., and .3-ma. long-delay groups. There was no evidence for any dynamogenic effect at any of the low shock intensities.

Figures 1 and 2 also indicate the difference in suppression which may be attributed to the temporal relationship between response and punishment. In most cases the difference was small, and there was considerable overlap between the contingent and long-delay *S*s at a given intensity. Taken individually, only two of the five comparisons of the mean suppression ratio during punishment training reached statistical reliability ($U = 4$, $p < .001$ for the .2-ma. group and $U = 13$, $p < .025$ for

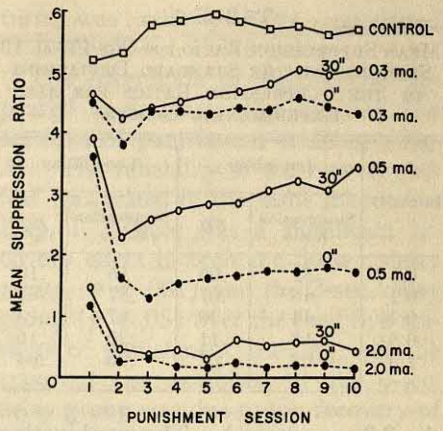


FIG. 1. Mean suppression ratio as a function of sessions for groups with 0.0-sec. and 30.0-sec. delay of punishment at intensities of .3, .5, 2.0 ma. and unpunished control group.

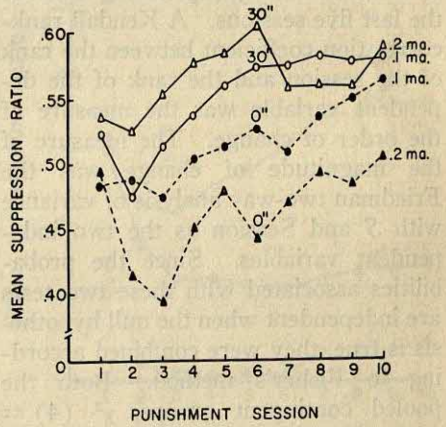


FIG. 2. Mean suppression ratio as a function of sessions for groups with 0.0-sec. and 30.0-sec. delay of punishment at intensities of .1 and .2 ma.

---

[4] All analyses were performed using nonparametric statistics because of heterogeneity of variances (see, e.g., Table 1).

TABLE 1

MEAN SUPPRESSION RATIO FOR THE FINAL 10 SESSIONS AND THE STANDARD DEVIATIONS OF THE SUPPRESSION RATIOS FOR ALL EXPERIMENTAL GROUPS

Intensity	Immediate		Long Delay	
	Suppression Ratio	*SD*	Suppression Ratio	*SD*
0.1	.52	.06	.56	.10
0.2	.46	.07	.57	.02
0.3	.44	.20	.47	.13
0.5	.18	.11	.30	.16
2.0	.03	.02	.06	.04

the 2.0-ma. group). The combination of the probabilities from the five independent *U* tests (Fisher's method) showed the overall delay effect to be highly reliable, χ^2 (10) = 30.9, $p <$ .001.

Relative to the unpunished control group the experimental groups appeared to be decreasingly affected by the punishment on successive sessions of training. In order to test the reliability of the recovery two statistical tests were combined to reflect both the trend and the magnitude of the daily differences between the mean suppression ratio of the experimental groups and the control group during each of the last five sessions. A Kendall rank-correlation coefficient between the rank of the session and the rank of the dependent variable was the measure of the order of change. The measure of the magnitude of change was the Friedman two-way analysis of variance with *S* and Session as the two independent variables. Since the probabilities associated with these two tests are independent when the null hypothesis is true, they were combined according to Fisher's method. Both the pooled contingent groups, χ^2 (4) = 10.0, $p < .05$, and the pooled noncontingent groups, χ^2 (4) = 16.1, $p < .01$, displayed a significant recovery over the final five punishment sessions.

Table 1 shows the variability of suppression ratio as a function of the intensity of the punishment for both the contingent and the long-delay groups. The variability of the punishment ratios was substantially greater in the middle of the intensity ranges than at either extreme. With relatively small number of *S*s per group, this variability may mask the effect of contingency of punishment particularly in the middle of the intensity range.

EXPERIMENT II

The absolute magnitude of the difference in response suppression between a contingent and a long-delay group at any particular intensity was small in the first experiment, although the phenomenon was not masked by an intensity of shock that was too extreme since the entire effective range of intensities was sampled. In the present experiment an interval, rather than a ratio, schedule of punishment was used (Camp, Raymond, & Church, 1966). One of the purposes of the present experiment was to establish a set of conditions that would result in a large absolute difference in the magnitude of response suppression between a contingent and a long-delay group.

Although the initial investigations of systematic delay of punishment did not lead to a clear picture of the temporal relationship between response and punishment (Bevan & Dukes, 1955; Warden & Diamond, 1931) more recent studies suggest that a delay of punishment gradient occurs in a wide variety of situations. Kamin (1959) found that the number of avoidance responses during extinction was directly related to the delay of punishment, and Karsh (1965) found that the speed of an appetitive response was directly related to the delay of punishment. Similarly, Baron (1965) found that speed of a runway response was a decreasing function of the delay of punishment. Banks and Vogel-Sprott (1965), using human *S*s, demonstrated the greater effectiveness of immediate punishment as compared to delays of 30 sec. and longer. The results of Exp. I suggest that there is also a delay-of-punishment gradient in the free-responding situation such that the amount of response suppression increases as the interval between response and punishment decreases. The present study investigates this relationship.

Experimental design.—The 45 *S*s in the present experiment were distributed equally among three conditions of delay of punishment, (*a*) 0-sec. delay (contingent), (*b*) 2-sec. delay, and (*c*) 30-sec. delay.

Procedure.—Following five sessions of variable-interval reinforcement training, *S*s received the punishment training procedures for the final 10 sessions. During punishment training reinforcement continued to be available on a 1-min. variable-interval schedule and, in addition, punishments were available on a 2-min. variable-interval schedule. When punishment became available an *S* in the 0-sec. delay group received a punishment immediately following its next response; an *S* in the 2-sec. delay group received a punishment 2.0 sec. following its next response; an *S* in the 30-sec. delay group received a punishment 30 sec. following its next response.

Results.—The mean suppression ratio over the 10 sessions of punishment training was calculated for each *S*, and Fig. 3 shows this measure for each of the three delay groups during the 10 sessions of punishment training. A Kruskal-Wallis analysis of variance on the mean suppression ratio during the 10 sessions of punishment training indicated that there was a significant effect due to delay, $H(2) = 39.6$, $p < .001$. Mann-Whitney *U* tests ($\alpha = .05$, 2-tailed) between all pairs of delay groups indicated that the 0-sec. delay group was reliably more suppressed than the 2-sec. delay group, which, in turn, was reliably more suppressed than the 30-sec. delay group.

The magnitude of suppression appeared to decrease after the first few sessions of punishment training (Fig. 3). The reliability of the recovery effect was tested in the same manner as Exp. I. There was a significant recovery effect in both the 30-sec. delay group ($p < .001$) and the 2-sec. delay group ($p < .05$) over the final five sessions of punishment training, but the apparent rise exhibited by the 0-sec. delay group was due to the recovery of only one *S*.

The variability of the mean suppression ratios over the 10 sessions of punishment training was not equal among conditions. The most suppressed group, contingent punishment, had the least variability ($SD = .05$). The 2-sec. and 30-sec. delay of punishment groups had greater variability ($SD = .17$ and $SD = .18$, respectively).

EXPERIMENT III

In both Exp. I and II there was greater response suppression by a con-

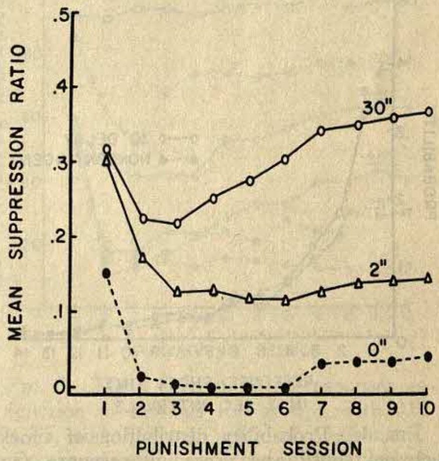


FIG. 3. Mean suppression ratio as a function of sessions for groups with 0.0, 2.0, and 30.0-sec. delay of punishment.

tingent punishment than by a punishment delayed by 30 sec. The use of a 30-sec. delayed punishment as representative of a "long" delay was based on the assumption that, in the free-responding situation, it would be essentially indistinguishable from a noncontingent aversive shock (one that is independent of *S*'s behavior). The present experiment justifies this assumption. The two procedures were directly compared with respect to (*a*) the actual treatment received, as measured by the frequency distribution of response-shock intervals, and (*b*) the resulting behavior, as measured by the suppression ratio. If the long-delay and noncontingent procedures are essentially indistinguishable, then the results of Exp. I and II can be related to previous studies comparing the contingent and noncontingent procedures. These studies have not resulted in consistent conclusions. For example, Estes (1944) found no apparent difference in resistance to extinction between a group of rats that received a brief period of punishment immediately following a response (Exp. B) and a group that received punishment independent of responding (Exp. I). On the other hand, Azrin (1956), using pigeons in a punishment training situation, found that contingent punishment resulted in a greater suppression than noncontingent punishment.

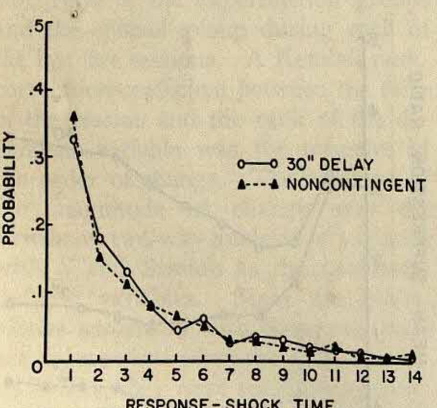


FIG. 4. Probability distribution of times between shock and preceding response (in 2-sec. intervals) for groups with noncontingent shock and 30-sec. delay of punishment.

Experimental design.—Thirty rats were randomly assigned to two groups of 15 *S*s each, (*a*) a noncontingent shock group, and (*b*) a 30-sec. delay-of-punishment group.

Procedure.—Following five sessions of variable-interval reinforcement training, *S*s received differential treatment during the next 20 sessions. The 1-min. variable-interval schedule of reinforcement continued in effect. On the average of once every 2 min. (*a*) the noncontingent group received a shock regardless of its behavior, and (*b*) the 30-sec. delay-of-punishment group received a shock 30 sec. after its next response. (Thus, an *S* in the noncontingent group received 15 shocks in each 30-min. session and an *S* in the 30-sec. delay group would receive 15 shocks in a session if it continued to respond more than once every minute, the minimum time between scheduled shocks.)

Results.—The probability distribution of time between shock and preceding response in the first 14 2-sec. intervals after a response is shown in Fig. 4. The functions for the two groups were virtually identical through this range although, of course, the two groups differed after that point (i.e., all the remaining response-shock intervals were approximately 30 sec. in the 30-sec. delay group but they could be of any length in the noncontingent group). Twenty-one percent of the response-shock intervals in the 30-sec. delay group and 24% of the response-shock intervals in the noncontingent group were longer than 28 sec.

The mean suppression ratios of the 30-sec. delay and the noncontingent groups were compared over the first 10 sessions, the last 10 sessions, and all 20 sessions. Mann-Whitney *U* tests indicated there were no significant differences in suppression ($U = 73$, $p > .10$;

$U = 85$, $p > .10$; $U = 75$, $p > .10$, respectively). No figure is provided of the mean suppression ratio over sessions for these two groups since they were similar to one another and both were similar to the 30-sec. delay group in Exp. II. The mean suppression ratio of the 30-sec. delay group in the present experiment was indistinguishable from the 30-sec. delay group in Exp. II ($U = 82$, $p > .10$) so that the results from these two experiments can be compared directly.

Both the 30-sec. delay-of-punishment and the noncontingent groups exhibited reliable recovery effects over Sessions 6 through 10 of punishment training ($p < .001$). (The method employed to test for recovery was the same as the one used in the previous experiments.) The variability of the 30-sec. delay group ($SD = .11$) was similar to that of the noncontingent group ($SD = .12$).

EXPERIMENT IV

In the free-responding situation, *E* may control the interval between a shock and the response that produces it, but the distribution of intervals between shock and preceding response is under control of *S*. To minimize the problem of intertrial responding, the present experiment is an investigation of the delay of punishment gradient in a situation in which *S* is trained to make a single response in the presence of a discriminative stimulus.

Experimental design.—Eighty-four *S*s were used in this experiment. Eighteen *S*s per group were used in the 0.0-, 7.5-, and 30-sec. delay groups, and 10 *S*s per group were used in the 2.0-sec. delay group, the noncontingent shock group, and the no-shock control group.

Procedure.—During the 10 sessions of discrimination training a clicking sound (S+) was scheduled to occur on the average of once each minute. If *S* made a response in the preceding 15 sec., the onset of S+ was delayed by 15 sec. A response in the presence of S+ resulted in (*a*) the presentation of a 45-mg. Noyes food pellet, and (*b*) the termination of the stimulus. During the last 5 sessions of discrimination training if no response occurred during the first 10 sec. of S+, the stimulus automatically terminated.

During the next 12 sessions (punishment training), punishments were integrated in the following manner. For all groups reinforcement occurred immediately after the response on an average of 50% of the trials. The treatment on the other 50% of the trials differed among the various groups. A punishment occurred either immediately after the response, 2.0 sec. after the response, 7.5 sec. after the response, or 30 sec. after the response (in the case of the 0.0-sec., 2.0-sec., 7.5-sec., and 30-sec. delay groups). A control group and a noncontingent-shock group received no punishment on these trials. On the average of once every 2 min. in the absence of S+ the noncontingent shock group received a shock irrespective of its response during S+.

Results.—Figure 5 shows the mean percentage of discriminative stimuli to which a response was made by sessions for the six groups. A Kruskal-Wallis analysis of variance on the mean percentage of responses during the 12 sessions of punishment training was calculated for the six groups, and it yielded

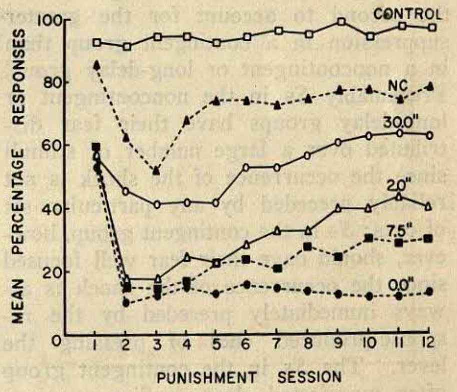


FIG. 5. Mean percentage of responses as a function of sessions for groups with 0.0, 2.0, 7.5, and 30.0 sec. delay of punishment, noncontingent shock, and unpunished control group. (Failure to press the lever within 10 sec. of stimulus onset was defined as nonresponse.)

an $H(5) = 43$, $p < .001$. Twelve of the 15 possible Mann-Whitney U tests between groups were significant at $p < .01$, and their results can be summarized as follows: (*a*) the control group made significantly more responses than any other group, (*b*) the noncontingent and the 30-sec. delay groups made significantly more responses than any group with a shorter delay, but they were not significantly different from each other, and (*c*) the group with immediate punishment (0.0-sec. delay) made significantly fewer responses than any other group.

DISCUSSION

The present experiments support the theoretical position that two mechanisms are necessary to account for the observed suppression in an experiment involving punishment of an appetitive response (*a*) suppression produced by the occurrence of an aversive stimulus and (*b*) suppression produced by the temporal relationship between a response and a punishment. The first mechanism is needed to account for the suppression of the noncontingent and long-delay groups relative to an unpunished control group, and the second to account for the greater suppression in a contingent group than in a noncontingent or long-delay group. Presumably *S*s in the noncontingent or long-delay groups have their fear distributed over a large number of stimuli since the occurrence of the shock is not reliably preceded by any particular set of cues; *S*s in the contingent group, however, should have their fear well focused since the occurrence of the shock is always immediately preceded by the response-produced cues of pressing the lever. The *S*s in the contingent group often manifested clear conflict reactions toward the lever. The greater suppression of lever responding in this group than in the noncontingent or long-delayed groups was probably due to the focusing of fear toward the lever, and this focused fear could have substituted for, rather than added to, the generalized anxiety of the noncontingent or long-delay groups.

The experiments involving the systematic delay procedure must be integrated into the framework proposed to account for the long-delay and contingent punishment results. Under the conditions of these studies the relatively long delay of 30-sec. can be considered equivalent to a noncontingent procedure. The shorter delay situation, however, can be viewed as a variation in the associative aspect of the situation or as a variation in the magnitude of the negative incentive. The *S* may be more likely to press a lever if punishment is delayed by 2 sec. than if it is immediate because (*a*) it is less certain that the punishment will occur, i.e., fear of a set of stimuli smaller than that of the noncontingent group but larger than that of the contingent group, or (*b*) it is less disturbed by a certain punishment that is not due for awhile than one that occurs immediately after a response.

Although further research is necessary to clarify the theoretical mechanisms, all the results of these experiments are consistent with the empirical generalization that the magnitude of suppression of an appetitive response is greater if the punishment is contingent upon the response than if it is not.

REFERENCES

AZRIN, N. H. Some effects of two intermittent schedules of immediate and nonimmediate punishment. *J. Psychol.*, 1956, 42, 3–21.

AZRIN, N. H. Effects of punishment intensity during variable-interval reinforcement. *J. exp. Anal. Behav.*, 1960, **3**, 123–142.

BANKS, R. K., & VOGEL-SPROTT, M. Effect of delayed punishment on an immediately rewarded response in humans. *J. exp. Psychol.*, 1965, **70**, 357–359.

BARON, A. Delayed punishment of a runway response. *J. comp. physiol. Psychol.*, 1965, **60**, 131–134.

BEVAN, W., & DUKES, W. Effectiveness of delayed punishment on learning performance when preceded by premonitory cues. *Psychol. Rep.*, 1955, **1**, 441–448.

CAMP, D. S., RAYMOND, G. A., & CHURCH, R. M. Response suppression as a function of the schedule of punishment. *Psychon. Sci.*, 1966, **5**, 23–24.

CHURCH, R. M. Systematic effect of random error in the yoked control design. *Psychol. Bull.*, 1964, **62**, 122–131.

CHURCH, R. M. The varied effects of punishment on behavior. *Psychol. Rev.*, 1964, **70**, 369–402.

ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57** (3, Whole No. 263).

HOFFMAN, H. S., & FLESHLER, M. Stimulus aspects of aversive controls: The effects of response contingent shock. *J. exp. Anal. Behav.*, 1965, **8**, 89–96.

HUNT, H. F., & BRADY, J. V. Some effects of punishment and intercurrent "anxiety" on a simple operant. *J. comp. physiol. Psychol.*, 1955, **48**, 305–310.

KAMIN, L. J. The delay-of-punishment gradient. *J. comp. physiol. Psychol.*, 1959, **52**, 434–437.

KARSH, E. Effects of number of rewarded trials and intensity of punishment on running speed. *J. comp. physiol. Psychol.*, 1962, **55**, 44–51.

KARSH, E. Delay of punishment at increasing intensity levels. Paper read at the Psychonomic Society, Chicago, October 1965.

WARDEN, C. J., & DIAMOND, S. A preliminary study of the effect of delayed punishment on learning in the white rat. *J. genet. Psychol.*, 1931, **39**, 455–461.

(Received May 27, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 124–131

PRIMARY STIMULUS GENERALIZATION OF THE GSR AS A FUNCTION OF OBJECTIVE AND SUBJECTIVE DEFINITION OF THE STIMULUS DIMENSION[1]

KENNETH R. BURSTEIN,[2] SEYMOUR EPSTEIN, AND BARRY SMITH

University of Massachusetts

2 experiments were conducted to determine whether the divergent results obtained in previous replications of Hovland's classic study of generalization to auditory frequency could be accounted for on the basis of Ss responding to a subjectively defined dimension rather than to the objective one defined by E. In 2 experiments Ss were asked to identify the test stimuli (tones of 153, 468, 1,000, and 1,967 cps). No support was obtained in either of the experiments for decremental gradients of stimulus generalization when response strength was plotted as a function of the perceived or physical characteristics of the stimuli. Accuracy of labeling the generalization stimuli as the CS produced highly reliable negatively accelerated decremental gradients.

Primary stimulus generalization has routinely been invoked as an explanatory construct by theorists in many disparate areas of psychology (viz., Mednick & Freedman, 1960). Although there has been disagreement about the precise form of the gradient, i.e., whether it is convex, concave, ogival, etc., it has generally been agreed that response strength falls off monotonically with increasing dissimilarity from the conditioned stimulus. While there are data to support such an assumption for operant conditioning (Guttman & Kalish, 1956; Honig, Boneau, Burstein, & Pennypacker, 1961), the data from classical conditioning paradigms leave the empirical status of decremental gradients in doubt. Hovland (1937a), Grant and Schiller (1953), and Wickens, Schroder, and Snide (1954) obtained monotonic decremental effects, Littman (1949) obtained results suggesting a zero slope and van Olst and Orlebeke (1965) and Epstein and Burstein (1966) found a U-shaped function.

In one of these studies (Epstein & Burstein, 1966), interviews following the experimental sessions revealed that some *S*s did not accurately perceive the stimulus conditions, i.e., they could not correctly identify the CS, the test stimuli, or the trials on which a UCS was presented.

Since it is possible that *S*s respond to their subjective environment rather than to the objective one constructed for them by *E*, it may be that the inability to obtain reliable gradients stems from the idiosyncratic perceptions of *S*s in the experimental situation.

The present studies, except for slight modifications, were replications of a previous study (Epstein & Burstein, 1966) which failed to obtain monotonic decremental gradients of stimulus generalization. One of the modifications consisted of asking *S*s to report on their perception of the stimuli presented to them, thus making it possible to analyze the data independently for subjectively and objectively defined stimulus dimensions.

---

[1] This paper, and the research reported in it, were supported by Grant MH 01293 given to S. Epstein from the National Institute of Mental Health, United States Public Health Service.

[2] Now at Simon Fraser University.

EXPERIMENT I

Method

Subjects.—The *S*s were 32 students, 16 males and 16 females, from introductory psychology courses at the University of Massachusetts. Six additional *S*s were discarded due to equipment failure or procedural error.

Procedure.—The *S* was seated in a reclining chair in a 6 × 6 ft. IAC soundproof room. The *E* and equipment were in an adjoining room. Communication between *S* and *E* was through intercom. After the shock and recording electrodes were attached and the instructions read, *S* was placed in the recording circuit and the shock intensity was set at a level which *S* reported as very unpleasant, but not painful. The stimuli were tones of 153, 468, 1,000, and 1,967 cps, the same frequencies employed by Hovland (1937). They were delivered by Eico sound generators through the two earphones of a matched Grason-Stadler headset. Equal subjective loudness levels were established in preliminary pilot work. Sound pressure readings at the earphones were 63, 70, 79, and 82 db., respectively. An electronic switch was introduced into the circuitry to eliminate clicks from the controlling relays. Following Hovland's procedure, the CS for half the *S*s was the 153-cps tone, and for the other half the 1,967-cps tone. The CS duration was 500 msec. The UCS was a shock delivered to the left wrist. The shock generator and electrodes are described in detail by Tursky and Watson (1964). The UCS duration was 100 msec. with its onset simultaneous with the offset of the CS, providing an interstimulus interval of 500 msec. The intertrial interval varied from 15 to 60 sec. with a mean of 37.5 sec. The sequence and duration of all events were automated by a series of electronic timers and relays.

The GSR was picked up on a Fels Dermohmmeter as a dc resistance change. Zinc electrodes, ⅞ in. in diameter, in a plastic cup filled with a zinc-sulphate electrolyte were clamped to *S*'s right and left palms. Signal marker pens recorded the onset and offset of stimulus events.

A programming module was designed which made available 24 Latin-square arrangements of four cycles of the four tones. For each *S*, one of these Latin squares was randomly selected to present the tones for 16 adaptation trials consisting of four cycles of four tones. Following the adaptation trials there were 16 conditioning trials, consisting of the 153-cps tone paired with the UCS for one-half the *S*s and the 1,967-cps tone paired with the UCS for the other half. Following the conditioning trials, another Latin square was selected to present 16 test trials, again consisting of four cycles and of four unpaired tones.

Instructions.—After being placed in the recording circuit, and before the adaptation trials were initiated, *S*s were presented with the four different tones in descending order, labeled as ABLE, BAKER, CHARLIE, and DOG. These labels were reversed for half the *S*s. To avoid the possibility of verbal reports influencing the response measure, *S*s were instructed to wait until after the termination of a tone before identifying the stimulus by calling out the label appropriate to it. After the 16 adaptation trials *S*s were instructed to listen to the tones in the next phase of the experiment without responding The *S*s then received 16 conditioning trials, after which they were again told to label the tones. Thus, *S*s labeled the tones during adaptation and test trials, but not during conditioning trials. The *S*s were asked to refrain from sudden or abrupt movements; they were instructed to relax, and were cautioned against falling asleep.

Results

Identical analyses of variance were performed on the pre- and postconditioning data. Any discernable deflection from the base line having a latency of 1.0 to 3.0 sec. was scored as a response. The actual measurement was accomplished with a Benson-Lehner Oscar K. A computer program transformed these measurements into conductance change in micromhos.

Preconditioning phase.—There were no significant effects attributable to the different tones ($F < 1$) or to the particular labels *S*s were instructed to use in identifying the tones ($F < 1$). Mean response strengths to the 1,967-, 1,000-, 468-, and 153-cps tones, respectively, were 1.23, 1.13, 1.25, and 1.36 micromhos. There was a significant adaptation effect as a function of successive cycles of tones, $F(3, 84) = 11.32$, $p < .001$. The Cycle × Tone interaction was not significant.

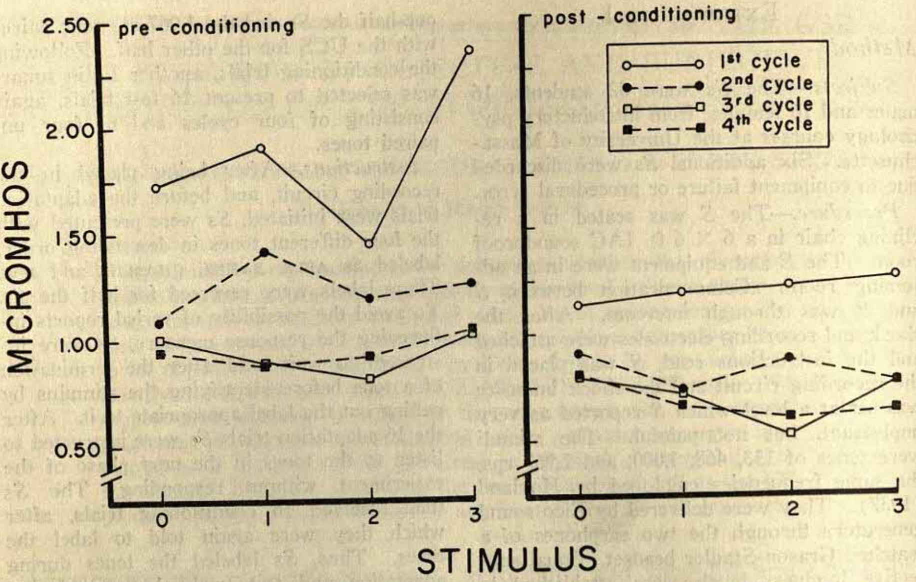


FIG. 1. Response strength for the pre- and postconditioning phases of Exp. I as a function of the objective stimulus dimension (CS = 0).

In order to facilitate comparison, the results for the preconditioning phase are plotted in Fig. 1 as a function of distance from the scheduled CS. No CS, of course, was presented in this phase of the experiment.

Postconditioning phase.—The results for the generalization phase of the experiment are presented in Fig. 1. There is no indication of a decremental gradient of stimulus generalization. The difference in response magnitude to the four stimuli is not significant, F (3, 84) = 1.09. The function obtained on the first cycle is similar to that reported by Littman (1949), being nonsignificantly different from zero slope. There is a highly significant difference in response strength as a function of test cycles, F (3, 84) = 11.26, $p < .01$, attesting to the rapid extinction of the response after the first test cycle. It is of interest that the level of response on the first cycle of the postconditioning phase is below that obtained on the first cycle of the preconditioning phase. When the data were pooled across cycles, the postconditioning response level was still lower than the preconditioning response level. However, response level on the first postconditioning cycle was higher than response level on the last cycle of the preconditioning phase, a finding which has frequently been interpreted as reflecting the occurrence of conditioning. The Tone × CS-Group interaction is highly significant, F (3, 84) = 8.27, $p < .001$, reflecting the markedly different response levels of the CS-groups to the stimulus farthest removed from the CS. The group which received the high frequency CS displayed its greatest response to the most remote CS, while the group which was conditioned to the low frequency CS exhibited its lowest response strength to the most remote CS.

Inspection of the data for individual *S*s on the first test cycle revealed that no *S* produced a decremental gradient of response strength, although three *S*s displayed a "reverse" gradient, i.e., increasing response strength with increasing remoteness from the CS.

Generalization was also evaluated by a between-*S* design in which only the first response of each *S* in the generalization test was used. The result of this analysis for the high frequency CS group was a U-shaped function with the largest response occurring to the test stimulus farthest removed from the CS. For the low frequency CS group the function was inverted-U shaped, with the largest response occurring to the 1,000-cps stimulus.

Figure 2 presents the data plotted on the stimulus dimension defined subjectively, i.e., as a function of *S*'s labeling of the stimulus. Neither of the CS subgroups exhibited a gradient when their data were plotted separately, although a very shallow gradient was obtained on the first cycle for the pooled CS subgroups. A separate analysis of this cycle revealed no significant effect ($F < 1$).

Mean response strengths were also plotted (*a*) using only stimuli correctly identified, i.e., stimulus and report were identical, (*b*) using only incorrectly identified stimuli and plotting GSR as a function of the objective stimulus, and (*c*) using only incorrectly identified stimuli and plotting GSR as a function of *S*'s report of the stimuli. No evidence was obtained for decremental gradients, or orderly gradients of any kind, under any of these conditions.

Each *S* was also assigned a score representing the number of errors in identification made in the generalization phase of the experiment. Two groups were formed on the basis of this score, a "good" discriminator group, consisting of the 10 *S*s who made the least (0–1) errors during the 16 test trials, and a "poor" discriminator group, consisting of the 10 *S*s who made the most (6–10) errors on the test trials. Analyses of variance revealed no significant differences in the response strengths of these two groups either in the preconditioning or postconditioning phases of the experiment.

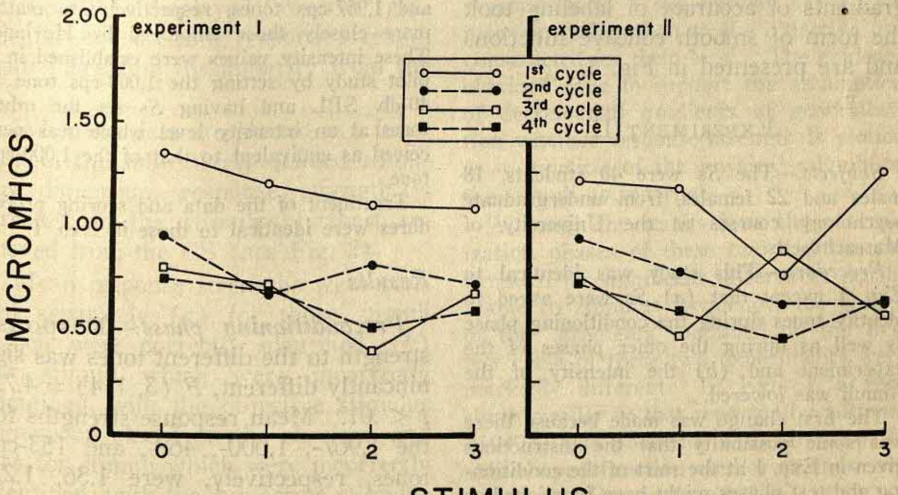


FIG. 2. Response strength for the pre- and postconditioning phases of Exp. I and II as a function of the subjective stimulus dimension (CS = 0).

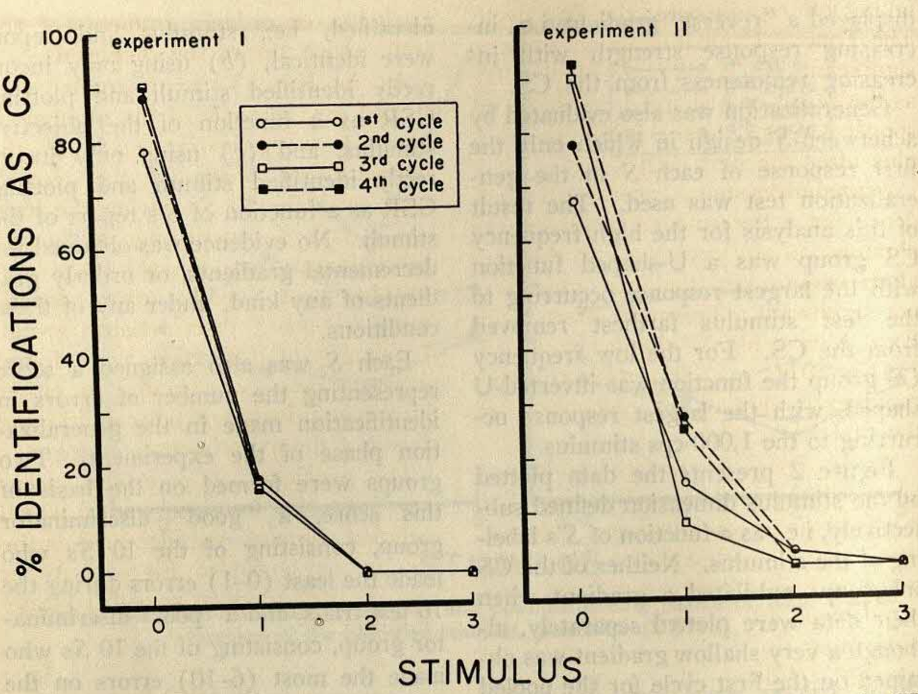


FIG. 3. Frequency of identification of stimuli as the CS as a function of distance from the CS.

In addition, the frequency with which each test stimulus was identified as the CS was plotted as a function of its distance from the CS. The resulting gradients of accuracy of labeling took the form of smooth concave functions and are presented in Fig. 3.

EXPERIMENT II

Subjects.—The *S*s were 40 students, 18 males and 22 females, from undergraduate psychology courses at the University of Massachusetts.

Procedure.—This study was identical to Exp. I except that (*a*) *S*s were asked to identify tones during the conditioning phase as well as during the other phases of the experiment and (*b*) the intensity of the stimuli was lowered.

The first change was made because there was some possibility that the instructions given in Exp. I at the start of the conditioning and test phases might have functioned as cues for the presence or absence of the UCS during these phases. It was also considered that labeling stimulus presentations during the conditioning trials would help to focus *S*'s attention on the CS.

In this experiment, sound pressure readings at the earphones were lowered to 70, 56, 40, and 54 db. for the 153-, 468-, 1,000-, and 1,967-cps tones, respectively, to match more closely those employed by Hovland. These intensity values were established in a pilot study by setting the 1,000-cps tone at 40-db. SPL and having *S*s set the other tones at an intensity level which was perceived as equivalent to that of the 1,000-cps tone.

Treatment of the data and scoring procedures were identical to those in Exp. I.

Results

Preconditioning phase.—Response strength to the different tones was significantly different, $F\ (3, 114) = 4.72$, $p < .01$. Mean response strengths for the 1,967-, 1,000-, 468-, and 153-cps tones, respectively, were 1.36, 1.25, 1.03, and 1.14 micromhos. There was a significant cycle effect, $F\ (3, 114) = 14.00$, $p < .001$, with response strength falling off rapidly after the first cycle. The Tone × Cycle interaction was also significant, F (3, 342) = 3.74, $p < .05$, as the result of a sharp decrease in response strength to the 1,967 cps stimulus after the first cycle.

Postconditioning phase.—No evidence for a decremental gradient was obtained on any cycle. As in the previous experiment, the tone effect was not significant. The cycle effect, as in Exp. I, was highly significant, F (3, 114) = 10.69, $p < .001$, reflecting the rapid extinction of the response after the first test cycle.

As in Exp. I, response level was lower subsequent to conditioning than it was prior to conditioning. This was true both when first cycles were compared and when performance was compared over the pooled cycles. Also confirming the findings in Exp. I, response level on the first postconditioning cycle exceeded that obtained on the last preconditioning cycle.

Inspection of the data from the first cycle of the postconditioning phase revealed that only 1 of the 40 *S*s displayed a decremental gradient. Two of the 40 *S*s displayed "reverse" gradients, i.e., increasing response strength with increasing dissimilarity from the CS.

When the data are plotted as a function of the subjectively defined stimulus dimension, response strength is greatest to the stimulus farthest removed from the CS (see Fig. 2).

Mean response strengths were plotted separately (*a*) for those stimuli which were correctly identified, (*b*) for stimuli which were incorrectly identified, with the objective stimulus dimension placed on the abscissa, and (*c*) for stimuli which were incorrectly identified, with the subjective stimulus dimension placed on the abscissa. No evidence was obtained for decremental gradients, or orderly gradients of any kind, under any of the conditions.

As in Exp. I, two groups were formed on the basis of the number of errors in identification made in the generalization phase of the experiment, with the 10 *S*s having the least (0–2) errors constituting a "good" discriminator group and the 10 *S*s having the most (6–13) errors constituting a "poor" discriminator group. No significant differences were found between these two groups in either the preconditioning or postconditioning phases of the experiment.

The gradients of perceptual accuracy, i.e., the frequency with which each test stimulus was identified as the CS, are presented in Fig. 3. As in Exp. I, they produce smooth concave gradients.

Discussion

The purpose of these studies was to determine whether the widely divergent results obtained by previous *E*s who attempted to repeat Hovland's classic study of generalization to auditory frequency could be accounted for on the basis of *S*s responding to a subjectively defined dimension rather than to the objective one constructed for them by *E*. The evidence obtained fails to support the assumption of decremental gradients of generalization whether response strength is plotted as a function of the physical stimuli or the stimuli as reported.

Although the results from the generalization phases of these two experiments revealed no significant differences attributable to stimulus frequency, the shapes of the "gradients" obtained on the first test cycle of these two experiments were markedly different. In Exp. I, a zero slope, similar to that reported by Littman (1949), was obtained, while in Exp. II a U-shaped function similar to those reported in Hovland's (1937b) study of intensity generalization, and in studies of frequency generalization by van Olst and Orlebeke (1965), and Epstein and Burstein (1966) was obtained.

It is quite possible that the elusiveness of conditioned generalization gradients in humans is directly related to the use of the galvanic skin response, the orienting component of which may be so sensitive as to obscure those aspects of the response which are a product of conditioning. Certainly GSR conditioning is somewhat unique in that, in the strictest sense of the term, there is no single class of stimuli which can serve as a specific conditioned stimulus for the GSR, or, putting it otherwise, there is no stimulus which does not initially elicit the response that is to be conditioned. Because it is quite difficult, and perhaps impossible, to adapt out the orienting component completely, and because it is, at present, equally impossible to determine quantitatively what proportion of a unitary response represents unconditioned sensory arousal and what proportion reflects the conditioned response, the use of this response may be inappropriate when the phenomenon under investigation requires analysis of slight differences in response magnitude (Gray, 1965).

Aside from the problem of being contaminated by an orienting component, GSR conditioning has another unique characteristic which may be related to the difficulty in producing the expected generalization gradients. Conditioning procedures involving the GSR typically generate acquisition curves which are quite different from those obtained with other classically conditioned responses, e.g., the eyelid response. Thus, Martin (1963) commented that "the general pattern of these acquisition curves is not that which would be expected on the basis of increasing habit strength," and several other studies have failed to obtain the negatively accelerated asymptotic performance typically associated with conditioning procedures (Bierbaum, 1958; Fjeld, Stern, & Zimny, 1964; Moeller, 1954; Stewart, Stern, Winokur, & Fredman, 1961).

Since conditioning is a prerequisite for the occurrence of generalization, and since GSR conditioning procedures result in atypical acquisition curves, it is conceivable that the procedures employed in Hovland-like experiments do not permit conditioning of the GSR. Supporting the possibility that the failure to obtain generalization gradients may be due to the absence of conditioning is in the finding in these studies that the mean GSR level in the postconditioning phase was lower than that obtained in the preconditioning phase. Unfortunately, the .5-sec. interstimulus interval typically employed in Hovland-like experiments does not permit monitoring the development of the conditioned response.

On the possibility that gradients might be obtained if only *S*s were used whose responses gave some obvious indication of conditioning, response strengths were plotted using only those *S*s who showed a greater response magnitude to the first test presentation of the test stimulus than to the same stimulus when first presented in the preconditioning trials. Five such *S*s were obtained in Exp. I, none of whom produced a decremental gradient. The pooled data from these five *S*s produced a U-shaped curve. In Exp. II, six *S*s met the criterion. While one of these *S*s produced a decremental gradient, the function obtained for the pooled data was again U-shaped.

Lastly, it is worth noting that plotting the frequency with which each test stimulus was verbally labeled as the CS as a function of its distance from the CS produced monotonic gradients which were, under all conditions, orderly and reliable. While errors in discrimination, as demonstrated by responses other than the GSR, e.g., verbal labeling, may be a ***necessary*** condition, they are clearly not a ***sufficient*** condition for producing GSR gradients following repeated pairings of a CS with a noxious stimulus.

To summarize, whether the data were plotted on an objective or subjective dimension, whether they were obtained from good or poor discriminators, whether they were subjected to within-*S*s or between-*S*s analysis, whether they were viewed as individual or group data, decremental gradients of generalization were not obtained.

Clearly, the conditions necessary to generate the kind of generalization gradients demanded by contemporary psychological theories are not yet well enough understood to allow them to be routinely demonstrated in the laboratory. At least this seems to be the current state of affairs with respect to classical GSR conditioning with human *S*s.

REFERENCES

BIERBAUM, W. B. The temporal gradient in GSR conditioning. *J. gen. Psychol.,* 1958, **59,** 97–99.

EPSTEIN, S., & BURSTEIN, K. R. A replication of Hovland's study of generalization to frequencies of tone. *J. exp. Psychol.,* 1966, **72,** 782–784.

FJELD, S. P., STERN, J. A., & ZIMNY, G. H. Effects of CS-UCS interval upon conditioning and extinction of GSR. *J. psychosom. Res.,* 1964, **3,** 45–48.

GRANT, B. A., & SCHILLER, J. J. Generalization of the galvanic skin response to visual stimuli. *J. exp. Psychol.,* 1953, **46,** 309–313.

GRAY, J. A. Stimulus intensity dynamism. *Psychol. Bull.,* 1965, **63,** 180–196.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.,* 1956, **51,** 79–88.

HONIG, W. K., BONEAU, C. A., BURSTEIN, K. R., & PENNYPACKER, H. S. Positive and negative generalization gradients obtained after equivalent training conditions. *J. comp. physiol. Psychol.,* 1963, **56,** 111–116.

HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.,* 1937, **17,** 125–148. (a)

HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.,* 1937, **51,** 279–291. (b)

LITTMAN, R. A. Conditioned generalization of the galvanic skin reaction to tones. *J. exp. Psychol.,* 1949, **39,** 868–882.

MARTIN, I. A further attempt at delayed GSR conditioning. *Brit. J. Psychol.,* 1963, **54,** 359–368.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.,* 1960, **57,** 169–200.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.,* 1954, **48,** 162–166.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.,* 1961, **68,** 60–67.

TURSKY, B., & WATSON, P. O. Controlled physical and subjective intensities of electric shock. *Psychophysiology,* 1964, **1,** 151–162.

VAN OLST, E. H., & ORLEBEKE, J. F. The role of the orienting reflex in the generalization of a conditioned response. *Canad. J. Psychol.,* 1965, **19,** 56–60.

WICKENS, D. D., SCHRODER, H. M., & SNIDE, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.,* 1954, **47,** 52–56.

(Received June 3, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 132–136

GROUPING STRATEGIES IN DICHOTIC LISTENING: THE EFFECTS OF INSTRUCTIONS, RATE, AND EAR ASYMMETRY

WAYNE H. BARTZ,[1] PAUL SATZ, AND EILEEN FENNELL

University of Florida

Grouping strategies and ear asymmetry (EA) were investigated in the Yntema and Trask (1963) dichotic listening (DL) paradigm. Trials consisting of 3 word-digit pairs were recorded at 2 rates, 2 pr. and 1 pr/sec. 8 groups of 20 right-handed Ss differed in terms of rate and instructions for recall: Pairs, Types, Ears, and Free. Results showed that Types and Ears strategies did not differ at either rate, the Ear strategy was most frequent under free recall, and EA effects were significant. Results indicate significant EA effects in DL experiments and should be a factor in models of DL.

Yntema & Trask (1963) have recently proposed a model of short-term memory to account for dichotic listening findings. This model is in contrast to Broadbent's (1957, 1958) original model which hypothesized that dichotically presented information enters the memory system via separate sensory channels which operate one at a time. According to Yntema and Trask (1963), information is processed serially by a "processor" and stored in "memory" with tags which identify the input. The emphasis of this model is that recall is a search process: the processor searches tags in memory and retrieves them when located. Thus Yntema and Trask argue that it is not separate sensory channels which account for the tendency of *S*s to group information by ear (ear order of report: EOR) when digits are presented at very rapid rates. Rather, input is "tagged" as having been presented to the right or left ears and this is the principle of search in recall. They suggest that digits are recalled more efficiently by ear (EOR) since no other characteristic so neatly divides them into two groups. Yntema and Trask (1963) hypothesized that if stimulus materials were digit-word pairs, *S*s would be provided with a second, less ambiguous set of tags for grouping types of items. Using an instructed condition of recall and a rapid rate of presentation, results were consistent with the above hypothesis: Recall by types was superior to recall by the EOR.

A factor notably absent in models of dichotic listening is that of ear asymmetry: The right ear has been shown to be superior to the left in the recall of digits (Satz, Achenbach, Pattishall, & Fennell, 1965) and words (Bartz, Satz, Fennell, & Lally, 1967; Borkowski, Spreen, & Stutz, 1965). Satz et al. (1965) and Bartz et al. (1967) have suggested that the superiority of the EOR was due, in part, to a built-in asymmetry between ears for verbal materials under dichotic stimulation. No such mechanism has been reported for the Types strategy. Further, the underlying ear asymmetry has been shown to be strong enough to prevent temporal recall between word pairs even when the pairs are composed of meaningful word associations (Bartz et al., 1967).

---

[1] Now at Iowa State University.

The present research was designed to investigate the Yntema and Trask (1963) experimental paradigm and analyze for the effects of ear asymmetry. This research included a free-recall condition in addition to instructed recall conditions so that the frequency of the EOR and Types recall strategies could be compared. A free-recall procedure would replicate also the Broadbent paradigm in dichotic listening. Finally, in addition to the standard rate (1 pr/$\frac{1}{2}$ sec), this paradigm was replicated at a slow rate (1 pr/sec) in order to determine the efficacy of the Types strategy when more time is allowed for switching channels.

PROCEDURE

Subjects.—The *S*s in this experiment were volunteers enrolled in undergraduate psychology courses at the University of Florida where participation in experiments was a requirement. The *S*s served in this experiment only if they reported that they (*a*) had no history of auditory impairment, (*b*) were right-handed, and (*c*) were without experience in dichotic listening.

Auditory stimulation procedure.—The equipment consisted of a stereophonic tape recorder (Wollensak, Model 1580T) and a set of stereophonic earphones (Koss, Model Pro-4). Stimulus materials were recorded on separate channels in the following manner: First, all stimulus materials to be presented in one channel were recorded with an intraword interval of $\frac{1}{2}$ sec. or 1 sec. between the onset of the words and an intertrial interval of 10 sec. to allow for *S*'s report. This channel was then replayed and all stimulus materials for the other channel were recorded and matched with the material in the first for time of onset of stimulus items, rate, volume, and intertrial interval. The same voice was heard in both channels (author EF).

Stimulus materials.—The stimulus materials were the digits 0 (sounded "oh") through 10, omitting 7, and the 10 words employed by Yntema and Trask: BET, COIL, GOOD, HOUSE, JACK, MUFF, PART, ROAM, and WORSE. All the words were thus monosyllabic.

Trial types.—Each trial consisted of three word-digit dichotic pairs selected at random. Four types of trials were created as in the Yntema and Trask study: *Type 1* in which digits were presented to one ear, words to the other; *Type 2* in which the first pair was reversed so that digit-word-word was recorded in one channel and word-digit-digit in the other; *Type 3* in which the last pair was reversed; and *Type 4* where the middle or second word-digit pair was reversed. Thus recall by types of items would require switching channels once in Trial Types 2 and 3, twice 4, and none at all in Type 1.

Instructions.—All *S*s read typewritten instructions prior to listening. The instructions informed *S* as to the nature of dichotic listening and listed the words and digits that would be presented in the first of three practice trials. At this point, the printed instructions differed in four ways which defined the groups in this study. *Free recall* required *S* to simply recall as many of the items as possible. *Types* instructions required that *S* recall first all the words and then all the digits or vice versa. *Ear* or EOR instructions required that *S* first recall the items presented to one ear and then those presented to the other. Finally, the *Pairs* instructions required that *S* recall the dichotic pairs in temporal order.

Design and procedure.—The 160 *S*s were randomly assigned to the eight Instruction × Rate conditions, 20 *S*s to each group. Half of the *S*s in each group listened with the earphones in one position and the remaining half with the earphones in the opposite position. Each *S*, tested individually, listened to 3 practice trials (not scored) and then to 16 experimental trials, 4 each of the four trial types in random order. Trials were separated by a 10-sec. recall interval. The *S* recalled aloud and *E* recorded *S*'s responses by marking the order of recall on a mimeographed copy of the experimental trials.

Scoring and analysis.—The responses of the *S*s were scored as correct only if pronounced correctly. As in the Yntema and Trask experiment, recall under each of the instructional conditions was scored according to those instructions. Under Types instructions, only the first and second groups of types of materials were scored correct. That is, if *S* recalled two digits, one word, then a digit, the latter digit was scored as incorrect. The same scoring procedure was employed for the Ears instructional condition except that the groups of items had to be from first one ear and then the other.

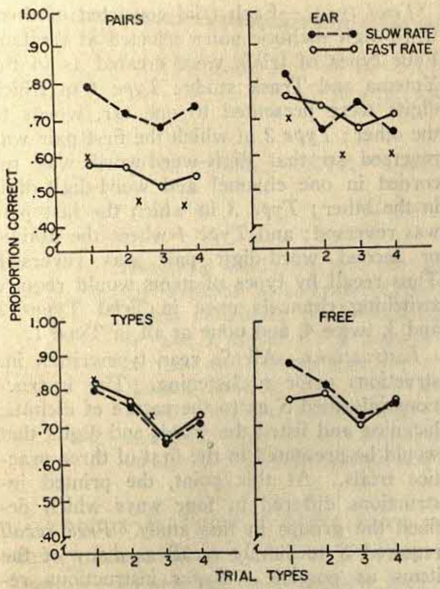


FIG. 1. Proportion correct recall for each experimental condition. (Additional data points represent the approximate fast-rate findings of Yntema & Trask, 1963.)

Under the Pairs condition, responses were correct only if *S* recalled the stimulus items by pairs. Finally, under free recall, any response was correct if pronounced correctly.

The number of stimulus items correctly recalled was analyzed by means of analysis of variance. A four-factor analysis was employed in which Instructions and Rate of presentation were between-*S* factors and Trial Types and Recall for items presented to each ear were analyzed as within-*S* factors (Winer, 1962). To assess order of report, a second analysis of variance was undertaken on the mean number of items correctly recalled for each ear as the primary and delayed channels. The first item recalled defined the primary channel; trials in which the first item was incorrect were omitted from the analysis. This second analysis of variance was identical to the first except that Ear and Ear Reported First were analyzed as within *S* factors.

Under Free recall, each trial was classified as the EOR, Types, Pairs, or Other strategy. A trial was classified as EOR if the first three items recalled had been presented to one of the ears, as Types if the first three items recalled were either words or digits, and as Pairs if the first four items recalled were dichotic pairs. Trials which did not conform to these criteria were classified as the "Other" strategy.

RESULTS

The results of the experiment are presented in Fig. 1 in terms of proportion correct recall. The additional data points represent the approximate fast-rate findings of Yntema and Trask (1963). The *S*s under the Pairs condition recalled significantly fewer items at the fast rate than did those under Pairs at the slow rate, $F\ (1,\ 152) = 16.42$, $p < .01$. In contrast to the Yntema and Trask findings where Types instructions resulted in greater recall than Ears instructions, no other difference among instruction-rate groups was significant ($p < .01$).

Trial types.—The main effect of Trial Types was significant, $F\ (3,\ 456) = 22.09$, $p < .01$. Tests among these means by the Newman-Keuls procedure (Winer, 1962) showed that the mean for Type 1 trials was significantly greater than the other means, and divergent with the results of Yntema and Trask, the Trial Type 3 mean was significantly lower than all others.

Trial Types were found to interact with Instructions, $F\ (9,\ 456) = 6.62$, $p < .01$. Tests by the Newman-Keuls procedure showed that the Trial Type means under each instructional condition did not differ at the slow rate of presentation. Significant differences were found only at the rapid rate: (*a*) there was lower mean recall for all Trial Type means under the Pairs condition, and (*b*) Type 1 trial means under Ears and Free recall were significantly lower than all Type 1 trial means at the slow rate ($p < .01$.)

Ear asymmetry.—Mean recall for items presented to the right ear was significantly greater than that for the left ear, $F\ (1,\ 152) = 28.87$, $p < .01$. The Ears × Instructions × Rate interaction was significant, F (3, 152) = 7.59, $p < .01$, indicating that the ear asymmetry was not present in all groups. Tests on the effects of ear asymmetry within each of the groups showed that the right ear was superior to the left in the following groups: At the slow rate, the Ears and Types groups ($p < .01$); at the fast rate, the Types group ($p < .01$) and also the Pairs and Free groups ($p < .05$).

Ear asymmetry and order of report.—Table 1 presents the mean recall for each ear when employed as the primary and delayed channel. This interaction of Ear and Order of Report was significant, F (1, 152) = 63.17, $p < .01$. Tests among these means by the Newman-Keuls procedure showed that the ear reported first (primary channel) was superior in recall to that reported second ($p < .01$). Mean recall for the ears did not differ significantly when employed as the primary or delayed channel ($p < .01$), however, the main effect of ear asymmetry was significant, F (1, 152) = 22.62, $p < .01$, with greater mean recall found in the right ear.

Recall strategies under free recall.—Table 2 presents the frequency percentages of each recall strategy that occurred under the Free-recall condition at both rates of presentation. The EOR was the most frequent recall strategy; it was identical to the Types strategy in Type 1 trials where the EOR occurred in two-thirds or more of the trials. The EOR was also most frequent in the other Trial Types which supposedly favor grouping by types of items. While the Types strategy was most frequent in Type 4 trials, this recall strategy is identical to what Bryden (1962, 1964) has termed the "attempted ear order of report." As in other investigations (Bryden, 1962, 1964), recall by temporal pairs was nonexistent at the fast rate, but occurred in approximately 8% of the trials at 1 pr/sec.

TABLE 1

EAR ASYMMETRY IN ORDER OF REPORT

Ear	Ear Reported First		Total
	Right	Left	
Right	2.43	1.97	2.20
Left	1.80	2.31	2.05
Total	2.12	2.14	

TABLE 2

PERCENTAGE OF RECALL STRATEGIES UNDER FREE RECALL

Recall Strategy	Trial Types							
	Slow Rate				Fast Rate			
	1	2	3	4	1	2	3	4
EOR	75[a]	50	54	38	66[a]	51	51	54
Types		15	9	16		10	15	21
Pairs	8	11	4	10	0	0	0	0
Other	17	24	33	36	34	39	34	25

[a] EOR and types strategies indistinguishable in Type 1 trials.

DISCUSSION

The present findings were not entirely consistent with the search model of Yntema and Trask (1963). This research found the right ear to be superior to the left in recall; ear asymmetry was also found in overall recall when order of report was taken into account. In addition, whereas Yntema and Trask predicted and found that the Type 4 trial (two crossings if recalled by types) resulted in lower recall than the other trial types, they grouped Types 2 and 3 (one crossing) together in their analysis. Keeping the Trial Types separate in this analysis showed that the Type 3 (one crossing after the second pair), not Type 4, resulted in lower mean recall than that for the other trial types.

The Yntema and Trask contention that a "types" tag was more reliable than an "ear" tag was not supported by this research. First, under instructed recall at the standard rate (1 pr/½ sec), the Types strategy was not superior to the EOR strategy. In terms of mean recall, both strategies were equally efficient and superior to the Pairs strategy, even though the stimulus materials (digits and words) supposedly favored grouping by types. Second, at the slow rate (1 pr/sec), the Types and EOR strategies again did not differ. One would expect, under the search model, a further bias in favor of the Types strategy because of the increased time available for searching tagged items across channels. The finding that the Pairs strategy was not different from the other strategies at the slow rate indicates that increased time was available; recall by pairs requires more channel crossing than does recall by types. Third, under Free recall, the EOR was the most frequent recall strategy, occurring in more than half the trials at both rates of presentation, again, even though the stimulus materials supposedly favored grouping by types.

The ear asymmetry findings have several implications. Methodologically, they point to the necessity of analyzing for ear asymmetry in dichotic listening experiments. Theoretically, the ear asymmetry findings: (*a*) support the Broadbent (1957, 1958) assumption of separate sensory channels, (*b*) show that ear asymmetry is a potent factor in the Yntema and Trask experimental paradigm, and (*c*) suggest that ear asymmetry is a factor which must be accounted for by models of dichotic listening. While the present authors have suggested elsewhere (Bartz et al., 1967; Satz et al., 1965) a rapproachement of the ear asymmetry findings and the Broadbent model, ear asymmetry could be accounted for in the Yntema and Trask model with the postulation of separate sensory channels. The present findings do not necessarily invalidate the Yntema and Trask hypothesis that recall is a search process, but do suggest that the search might best be considered as taking place within separate sensory channels.

REFERENCES

Bartz, W. H., Satz, P., Fennell, E., & Lally, J. R. Meaningfulness and laterality in dichotic listening. *J. exp. Psychol.*, 1967, **73**, 204–210.

Borkowski, J. G., Spreen, O. & Stutz, J. Z. Ear preference and abstractness in dichotic listening. *Psychon. Sci.*, 1965, **3**, 547–548.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957, **64**, 205–215.

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Bryden, M. P. Order of report in dichotic listening. *Canad. J. Psychol.*, 1962, **16**, 291–299.

Bryden, M. P. The manipulation of strategies of report in dichotic listening. *Canad. J. Psychol.*, 1964, **18**, 126–137.

Satz, P., Achenbach, K., Pattishall, E., & Fennell, E. Order of report, ear asymmetry and handedness in dichotic listening. *Cortex*, 1965, **1**, 377–396.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw Hill, 1962.

Yntema, D. B., & Trask, F. P. Recall as a search process. *J. verbal Learn. verbal Behav.*, 1963, **2**, 65–74.

(Received June 6, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 137–143

TRANSPOSITIONS IN SHORT-TERM MEMORY[1]

BENNET B. MURDOCK, JR. AND WALTER VOM SAAL[2]
University of Toronto *Columbia University*

This study attempted to determine whether transpositions occur during storage or during retrieval. A distractor task (Peterson & Peterson technique) was used in which word trigrams were followed by a variable-length retention interval. Words were used to minimize acoustic confusions, and the trigrams were composed of same-category or different-category items (e.g., GRAY BROWN RED vs. EEL SCREW LOUNGE). The same-category trigrams were better retained than the different-category trigrams but also transposed more. In addition, transpositions increased in frequency over the retention intervals used. Conrad has suggested that transpositions do not occur during storage; the results obtained here suggest the opposite conclusion.

Transpositions are a specific manifestation of a more general problem, that of the serial order of behavior. The importance of this general problem has been pointed out by Lashley (1951), Milner (1961), Penfield and Perot (1963), Sperry (1952), and many others. In the area of short-term memory (STM), Conrad (1959, 1965) has proposed an explanation for transpositions. He suggested that order and item information are not separate, that transpositions reflect acoustic confusability, and that reordering occurs at output not in storage. More specifically, he proposed a model wherein items at presentation were deposited in successive boxes; "During recall, *S*s 'read' the name of the item in each box in an order determined by the experimental instructions. They may start anywhere in the sequence . . . but there is no way in which the contents of the boxes can interchange [Conrad, 1965, p. 168]." Conrad has studied transpositions in a memory-span task; transpositions will be further studied here by means of a "distractor" technique, a term suggested elsewhere (Murdock, 1967) for the methods first used by Brown (1958) and Peterson and Peterson (1959).

As Conrad has found, transpositions are rare events. With the trigram ABC, a transposition may be defined as any one of the six permutations of ABC other than itself; i.e., ACB, BAC, BCA, CAB, CBA. In a previous experiment with trigrams (Murdock, 1961) the probability of a transposition was .023 whereas in a similar but later experiment (Murdock, 1967) it was .057; with 48 trials/*S* the mean values were 1.1 and 2.7 transpositions/*S*, respectively. In both cases (with *N*s of 24 and 48 *S*s, respectively) the obtained distribution across *S*s did not differ significantly ($p > .50$) from a Poisson with the means given above (i.e., 1.1 and 2.7).

The six possible permutations of ABC are shown in Fig. 1. The second two can be distinguished from the bottom three; that is, ACB and BAC are *adjacent* transpositions in that the place change involved is between contiguous

---

[1] This research was supported by Grant APA 146 from the National Research Council of Canada and by Grant GB 4545 from the National Science Foundation. The experiment was designed and conducted by WvS under the supervision of BM; the latter is responsible for the report and the interpretation of the data.

[2] Now at McMaster University.

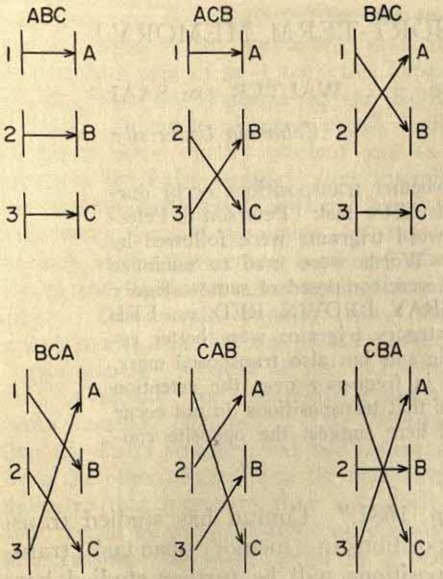


FIG. 1. Diagrammatic representation of correct responses (ABC), adjacent transpositions (ACB, BAC), and remote transpositions (BCA, CAB, CBA).

items; BCA, CAB, and CBA are *remote* transpositions in that the place changes involve more than adjacent items. The reason for distinguishing between adjacent and remote transpositions is as follows: As noted above, transpositions are rare events. By any reasonable mechanism one can envisage, adjacent transpositions should be more numerous than remote transpositions. Therefore, to distinguish between signal and noise (i.e., to insure that one is dealing with a real effect) one can set as a criterion the requirement that, in the data, adjacent transpositions outnumber remote transpositions. While the comparison between the two types of transpositions may be of interest itself, the frequency difference can serve as a guarantee of a genuine serial-order effect.

With the conventional distractor technique a variable-length retention interval is generally used. It can be argued (see Milner, 1961) that, at the time of attempted recall, *S* scans the memory trace and reports its contents. If transpositions occur during storage, then it seems reasonable to assume that the relative frequency of transpositions should *increase* over the retention interval. That is, while the absolute frequency of the correct recall of the three letters will certainly decrease, the conditional probability of a transposition (probability of a transposition given that the Items A, B, and C are correctly recalled) should increase. On the other hand, if Conrad is right and transpositions occur at output, then there is no obvious reason why the conditional probabilities should change at all. That is, in his model the contents of the boxes are intact, and no place change is possible.

To find out whether transpositions increase over time the data from the two experiments mentioned above were reanalyzed and are shown in Table 1. The bottom two rows show the absolute frequency of correct responses (ABC), adjacent transpositions (ACB, BAC), and remote transpositions (BCA, CAB, CBA) summed over *S*s and retention intervals for the two experiments. Even without consideration of the fact that a null hypothesis would

TABLE 1

CONDITIONAL PROBABILITIES (MURDOCK, 1967) AND TOTALS FOR ALL PERMUTATIONS OF ABC

Retention Interval	ABC	ACB, BAC	BCA, CAB, CBA
0 sec.	.992	005	.003
3 sec.	.912	.070	.017
6 sec.	.955	.045	—
9 sec.	.820	.168	.012
12 sec.	.811	.155	.034
18 sec.	.806	.155	.039
Total[a]	530	24	3
Total[b]	1264	109	19

[a] Murdock, 1961.
[b] Murdock, 1967.

predict relative frequencies of 1:2:3 it is quite clear that adjacent transpositions outnumber remote transpositions. For conditional probabilities, the upper rows show that, for the later experiment (Murdock, 1967), transpositions of both types do increase over time. The data from the earlier experiment showed the same trend, but they were too fragmentary to be very convincing.

However, these data are not conclusive because of the problem of acoustic confusions. As Conrad (1965) has argued, it could be that, over time, the traces became more illegible (Brown, 1959) and substitution occurred on the basis of acoustic factors (Conrad, 1964). And since sampling without replacement was used in the trigram construction one confusion led to another. Any analysis of these data is complicated by the fact that the stimulus material was originally chosen on the basis of Witmer (1935) association-value norms, and acoustic confusability is not easy to analyze with unselected consonants. As Conrad (1965) has said, "What would be crucial is a variable that, in the defined case, could be shown to affect order of items differentially from the items themselves [p. 169]."

The present study will show that there is such a variable; namely, category similarity. The stimulus items of the trigrams were either words from the same category (e.g., GRAY BROWN RED) or words from different categories (e.g., EEL SCREW LOUNGE). A variable-length retention interval was used to track any changes in frequency of transpositions over time. As will be reported, the category variable differentially affected accuracy of recall and transpositions. Finally, the substitution of words for consonants as components of the trigram should minimize acoustic confusions. In our previous studies with paired associates consisting of common English words, acoustic confusions have been very infrequent. As one example (Exp. III of Murdock, 1963) we found only 152 acoustic confusions out of a possible 4,992 when a lax criterion for "acoustic" was used; moreover, infrequent though they were, they *decreased* in absolute number by 50% over the brief (10-sec.) retention intervals used. With such material, then, an interpretation of transpositions in terms of acoustic similarity becomes unlikely.

METHOD

Procedure.—The distractor technique of Peterson and Peterson (1959) was used in which word trigrams were followed by a variable-length retention interval. The retention intervals used were 3, 9, and 18 sec., and the distractor task was *E*-paced random addition. That is, *S*s were read two-digit numbers to which they had to add three and report the sum (*E*: "58" *S*: "61"), and *E* presented the next number while *S* was reporting the previous sum. In general *S* was pushed as fast as possible short of being unable to respond at all.

The word trigram was presented visually with a duration (controlled by a Hunter timer) of 1.3 sec. (The duration had been found adequate for one reading of the items.) The first two-digit number was read immediately upon offset of the trigram. The end of the interpolated task (retention interval) was signaled by a loud gong. The *S* was instructed to recall the three words in the correct order, and encouraged to guess if not sure. While there was no check on the possibility of an occasional perceptual error (i.e., misreading a word) this precaution was unnecessary since the main analysis will deal with conditional probabilities. The reason for not having such a check (say by having *S* read the words aloud) is that the offset of the presentation period and/or the start of the retention interval would then effectively change from *E* paced to *S* paced.

Materials.—The stimulus materials consisted of 144 monosyllabic nouns taken from the cultural norms of Cohen, Bousfield, and Whitmarsh (1957). In all 22 categories were represented, 20 categories with 6 words

TABLE 2

Same-Category	Different-Category
GRAY BROWN RED	EEL SCREW LOUNGE
SKULL THIGH WAIST	ALE BOMB STEAL
OWL DUCK LARK	YACHT BEAM FLOOD
SPEAR AXE CLUB	WINE COAT PIG
SLOOP TUG RAFT	BIRCH CORN PINK

each and 2 categories with 12 words each. The criteria used in selecting the words were as follows (and in this order): (*a*) Syllabic length; (*b*) Representativeness; (*c*) Category dominance. As to *a,* all polysyllabic words were rejected. As to *b,* the requirement was that the word in isolation would unambiguously suggest the proper category (e.g., ORANGE as a color was rejected). As to *c,* the low dominant were preferred; thus CHAIR, BED, COUCH, and DESK were not used as articles of furniture; ARM, LEG, HEAD, and FOOT were not used as part of the human body; CAT and DOG were not used as four-footed animals. Illustrative trigrams are shown in Table 2. The average frequency with which the 144 words appeared in the Connecticut norms was 24; the range was 1–97. Less than 10% of the words finally selected were among the three dominant numbers of their appropriate category, and the median rank order of the 144 words within their respective category was below 10.

Conditions.—Category similarity was manipulated as a within-*S*s variable. That is, each trigram consisted either of three words from the same category or of three words from different categories (see Table 2). There were 24 trigrams of each type. In assembling the latter trigrams (i.e., from different categories) random sampling without replacement was used with the further restriction of no duplication of *a* categories or *b* initial letters within a trigram. Further, for counterbalancing, two completely different sets were used, each for half the *S*s; words that composed same-category trigrams for half the *S*s were assigned to different-category trigrams for the other half, and vice versa.

In addition, retention interval was also a within-*S*s variable, so there were eight trigrams of each type (i.e., same-category or different-category) at each retention interval (i.e., 3, 9, or 18 sec.). Order of presentation of the $2 \times 3 = 6$ conditions was randomized over the full 48 trials subject to the restrictions that (*a*) in Trials 1–24 and Trials 25–48 an equal number of same- and different-category items appear, and (*b*) no successive trials contain words from the same category. Four different random orders were prepared, with one-quarter of the *S*s within each set assigned to each order. Finally, over the entire experiment each item occurred equally often at each retention interval.

In the instructions *S*s were fully told about the nature of the category variable. That is, they were told that some trigrams would be words from the same category, other trigrams would not. Further, they were told the retention interval would vary from 3 to 18 sec. Practice on the interpolated task (the *E*-paced random addition) preceded the start of the experiment proper.

Subjects.—The *S*s were 48 students of both sexes from the introductory psychology course at the University of Vermont who were fulfilling a course requirement. All *S*s were tested individually.

RESULTS

Probabilities for correct recall of the ordered trigram (i.e., ABC) were, for same-category trigrams .67, .46, and .43; for different-category trigrams .44, .28, and .21 for 3-, 9-, and 18-sec. retention intervals, respectively. Since the overall superiority of the same-category trigrams was shown by 44 of 48 *S*s there is little doubt as to its statistical significance ($p < 10^{-7}$ by a sign test). If one uses a free-recall criterion (the three words in any order) the probability values are raised slightly, and this effect was shown by 47 of 48 *S*s. Thus, there seems little

TABLE 3

CONDITIONAL PROBABILITIES FOR SAME- AND DIFFERENT-CATEGORY TRIGRAMS

Interval	ABC	ACB BAC	BCA, CAB, CBA
Same			
3 sec.	.902	.080	.017
9 sec.	.798	.166	.036
18 sec.	.788	.159	.053
Different			
3 sec.	.939	.055	.006
9 sec.	.885	.107	.008
18 sec.	.882	.097	.021

doubt that trigrams composed of words from the same category are better retained in a distractor task than words from three different categories.

There were over three times as many transpositions with same-category trigrams as with different-category trigrams (specifically, 117 to 36). Since the direction of this difference was the same for each type of transposition at each retention interval (a 14–0 split with one tie) there is little doubt as to its statistical significance. Moreover, in both cases adjacent transpositions outnumbered remote transpositions; the ratio was nearly 4:1 for same-category items and 8:1 for different-category items.

A detailed analysis is shown in Table 3, which gives the data for correct responses, adjacent transpositions, and remote transpositions separately for same- and different-category trigrams. As is quite clear, correct responses *decrease* and transpositions *increase* over the retention interval. Furthermore, if one makes a point by point comparison between the same-category data shown in Table 3 and the comparable values in Table 1 the agreement could not be much closer. (The reason for selecting same-category trigrams rather than different-category trigrams for this comparison is that CCC trigrams also consist of items from the same category; namely, letters.) It would seem then that the relative frequencies of both adjacent and remote transpositions increase over a filled retention interval, and that the factor of acoustic confusability is not critical.

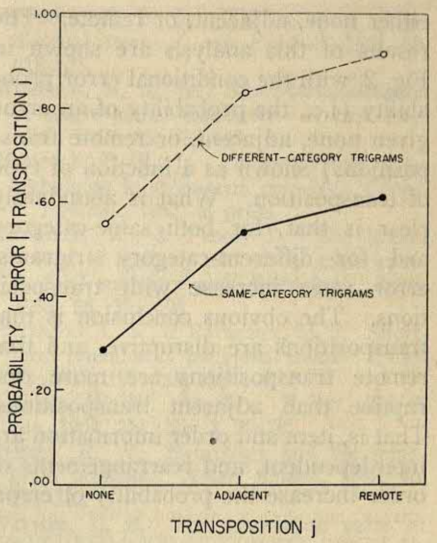


FIG. 2. Conditional probability curves for errors given none, adjacent, and remote transpositions.

In an attempt to shed further light on the processes involved a detailed analysis of the relationship between transpositions and errors was undertaken. Responses were classified into single errors (ABX, AXC, and XBC) and the five possible transpositions of each (for ABX the transpositions would be AXB, BAX, both adjacent, BXA, XAB, and XBA, all remote); also into double errors (AXX, XBX, and XXC) plus the two possible transpositions of each (for AXX the transpositions would be XAX, adjacent, and XXA, remote). By including also ABC (all correct) and XXX (all incorrect) this classification is exhaustive; all possible outcomes are included. After the tabulation was complete it was gratifying to find that the criterial relationship still obtained; for single and double errors adjacent transpositions were more numerous than remote transpositions both for same-category and for different-category trigrams. Furthermore, transpositions still increased over the duration of the retention interval.

Then, a two-way classification was used; for each category, errors (any noncorrect response) either were or were not present; transpositions were either none, adjacent, or remote. The results of this analysis are shown in Fig. 2, with the conditional error probability (i.e., the probability of an error given none, adjacent, or remote transpositions) shown as a function of type of transposition. What is abundantly clear is that, for both same-category and for different-category trigrams, error rates increase with transpositions. The obvious conclusion is that transpositions are disruptive, and that remote transpositions are more disruptive than adjacent transpositions. That is, item and order information are interdependent, and rearrangements of order increase the probability of error.

DISCUSSION

These results suggest that transpositions are lawful events which, while infrequent, can still be studied at the level of individual trigrams. Adjacent transpositions are more frequent than remote transpositions, and this relationship holds consistently across retention intervals. For the storage retrieval issue, transpositions do occur when there is little or no acoustic confusability, they increase in relative frequency over the retention interval, and the similarity variable showed that trigrams which are better remembered (same-category trigrams) transpose more frequently than do trigrams which are less well remembered (different-category trigrams). The opposite conclusion to that of Conrad (1965) is suggested; namely, that transpositions do in fact occur in storage.

The reason why the same-category trigrams were better retained than the different-category trigrams is certainly not clear. Initially the opposite had been expected, on the grounds of similarity effects and interference in STM. Given that the items used tended to be low-dominant members of the categories (see the third criteria for item selection) and that the categories used were nonexhaustive, an explanation in terms of guessing probabilities or recognition strategies seems unlikely.

It should be kept clearly in mind that Conrad (1965) used 6-gm. items with an immediate retention test whereas we used trigrams with delayed retention tests. In his study he reported 1,153 transpositions for 387 *S*s × 40 trials/*S* for an obtained proportion of .074; in the present study there were 153 transpositions for 48 *S*s × 48 trials/*S* for an obtained proportion of .066. However, if one disregards serial position effects it is reasonable to suggest that the base rate (chance level) should be proportional to n^2; hence the relative frequency of transpositions in the present study outnumbered those in the Conrad study by more than 4:1. Thus, to study transpositions, introduction of delay intervals appears to increase the amount of usable data even when acoustic confusability is minimized.

Finally, a more formal analysis of transpositions is possible. Six operations can be defined (the five "transpositions" plus the identity operation which leaves the order unchanged) and, in terms of modern algebra, these operations form a finite, noncommutative group. It is a group in that any two operations performed in succession are equivalent to another operation belonging to the set (e.g., ACB followed by BAC is equivalent to CAB), and noncommutative in that the order in which the operations are performed affects the outcome (e.g., BAC followed by ACB results in BCA, not CAB). Actually, this very example is illustrated in a recent article by Sawyer (1964), and it immediately suggests that the implicit assumption of only a single transposition may be unduly restrictive. If the conclusion of this article is generally true, then such a formal analysis may be useful in further studies of serial-order problems in memory.

REFERENCES

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12–21.

Brown, J. Information, redundancy and decay of the memory trace. In *The mechanisation of thought processes. Natl. Phys. Lab. Sympos.* No. 10. H.M.S.O., 1959.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

Conrad, R. Errors of immediate memory. *Brit. J. Psychol.,* 1959, **50**, 349–359.

Conrad, R. Acoustic confusions in immediate memory. *Brit. J. Psychol.,* 1964, **55**, 75–84.

Conrad, R. Order error in immediate recall of sequences. *J. verbal Learn. verbal Behav.,* 1965, **4**, 161–169.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

Milner, P. M. A neural mechanism for the immediate recall of sequences. *Kybernetik,* 1961, **1**, 76–81.

Murdock, B. B., Jr. The retention of individual items. *J. exp. Psychol.,* 1961, **62**, 618–625.

Murdock, B. B., Jr. Short-term retention of single paired associates. *J. exp. Psychol.,* 1963, **65**, 433–443.

Murdock, B. B., Jr. Distractor and probe techniques in short-term memory. *Canad. J. Psychol.,* 1967, in press.

Penfield, W., & Perot, P. The brain's record of auditory and visual experience: A final summary and discussion. *Brain,* 1963, **86**, 595–696.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual items. *J. exp. Psychol.,* 1959, **58**, 193–198.

Sawyer, W. W. Algebra. *Scient. American,* 1964, **211**, No. 3, 70–78.

Sperry, R. W. Neurology and the mind-brain problem. *Amer. Scient.,* 1952, **40**, 291–312.

Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.,* 1935, **47**, 337–359.

(Received June 10, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 144–146

SUPPLEMENTARY REPORTS

REPLICATION OF GSR AVOIDANCE CONDITIONING WITH CONCOMITANT EMG MEASUREMENT AND SUBJECTS MATCHED IN RESPONSIVITY AND CONDITIONABILITY[1]

H. D. KIMMEL AND H. S. STERNTHAL
Ohio University *University of Florida*

This study replicated Kimmel and Baxter's (1964) on avoidance conditioning of the GSR, with yoked pairs of avoidance and control Ss matched in responsivity and conditionability and with concomitant EMG measurement. The 19 pairs of Ss (of a total of 24) who were most closely matched in responsivity and conditionability differed in adjusted GSR magnitude during acquisition (but not quite significantly) and the difference vanished during extinction. It was concluded that the use of a yoked control design without matching Ss in responsivity and conditionability may tend to exaggerate the observed avoidance conditioning effect spuriously. The EMG measures were found to be unrelated to GSR responses and not capable of explaining the differences that were observed.

Recent studies (Kimmel & Baxter, 1964; Kimmel, Sternthal, & Strub, 1966) purporting to demonstrate avoidance GSR conditioning may be criticized on two grounds. Church (1964) has noted that the yoked design may favor the avoidance group over its yoked controls if pairs of *S*s are mismatched in initial conditionability. According to Church, when the avoidance *S* is more conditionable than the control, only a few UCSs will be needed for him to make CRs and avoid the UCS. His yoked control will, thus, not receive enough UCSs to achieve a high level of responding. But, when the yoked control is the more conditionable, the avoidance *S* will continue to receive the UCS until *his* level of responding results in avoidance. The yoked control will not benefit from the extra UCSs because he will have reached asymptotic performance.

The second difficulty is that the avoidance *S* may acquire a skeletal response instrumentally which may elicit unconditioned GSRs. The yoked control cannot learn a muscular response, since his behavior does not influence UCS occurrence. Thus, *apparent* GSR conditioning differences between the groups may be attributable to the muscular response.

The present study was an attempt to replicate the essential features of the earlier ones, at the same time matching the members of yoked pairs in both classical conditionability and responsivity. In addition, concurrent recording of EMG from *S*'s forearm was obtained to determine whether observed GSRs might have resulted from movements.

Method.—Forty-eight female undergraduates at the University of Florida volunteered to serve as *S*s. Grass EMG electrodes, covered with electrode cream, were attached to the upper surface of *S*s left arm, about two-thirds of the distance from the wrist to the elbow. The GSR was picked up from the palm and back of the left hand by ¾-in. zinc-zinc sulphate electrodes in lucite cups filled with NaCl jelly. Both responses were led into a Grass Model 5 polygraph (preamplifiers: 5P-1 for GSR and 5P-3 for EMG) with a paper speed of 5 mm/sec. The CS was a 1,000 cps 40-db. tone (physical reference) of 1.0 sec. duration, presented via earphones. The UCS was a 3.5-mamp. shock delivered to the right index and middle fingertips. Its duration was .1 sec. The CS-UCS interval was 5.0 sec.

The *S* was seated in an IAC audiometric chamber, illuminated by a 100-w. shielded lamp. Noninformative instructions were read over an intercom telling *S* to relax and pay attention. The first five *S*s to arrive were assigned to the avoidance group. Thereafter, *S*s matched in responsivity and conditionability were assigned as controls for previously run avoidance *S*s or, when no match was possible, were assigned to the avoidance group to await matching with some future *S*. In this way 24 pairs of matched *S*s were run. *Responsivity* was de-

---

[1] Done under United States Public Health Service Grant M-06060-02 and based in part on the second author's MA thesis at the University of Florida.

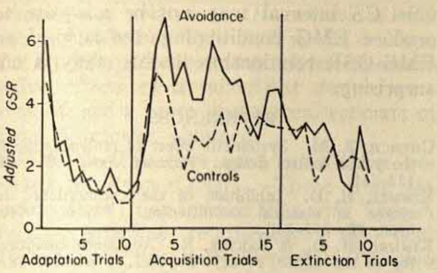


FIG. 1. Mean adjusted GSR magnitudes of best-matched avoidance and control groups ($N = 19$ in each) during adaptation, acquisition, and extinction (arbitrary units on the ordinate).

fined as the sum of the GSRs on the last of 10 adaptation trials and the first 2 acquisition trials. *Conditionability* was defined as the GSR on the second acquisition trial minus the GSR on the first acquisition trial. These criteria were used during the actual data collection. More accurate measures were used for the data analysis.

Following three initial trials of shock alone, *S*s received 10 presentations of the CS alone (adaptation). Then two paired CS-UCS trials were run. Fourteen additional acquisition trials were given, in which the avoidance *S* could prevent the occurrence of the shock both for himself and for his control by making an anticipatory GSR that equaled or exceeded the GSR on the second CS-UCS trial (avoidance was implemented manually by *E* who monitored the GSR record visually). Ten extinction trials followed. Then *S* was instructed to make various movements of his left hand and arm to ensure proper placement of the EMG electrodes.

Results.—The GSR measure used was the maximum change in log conductance following the CS but preceding the UCS (or 5 sec. on nonshocked trials). These were expressed as deviations from prediction from the multiple regression of the sum of the last 13 acquisition GSRs on responsivity and conditionability. Here *responsivity* was defined as the sum of the transformed GSRs on the last adaptation trial and the first three acquisition trials (shock could not be avoided until the third trial). *Conditionability* was defined as the sum of the transformed GSRs on the second and third acquisition trials (i.e., before pairing) minus the sum of the transformed GSRs on the last adaptation and first conditioning trials (i.e., after pairing).[2]

Since the groups tended to differ even before pairing, and since five pairs of *S*s had not been as well matched during data collection as the remaining 19, their data were set aside. The adjusted GSR measures of the 19 pairs of *S*s who were most closely matched in both responsivity and conditionability are presented in Fig. 1, for the adaptation, acquisition, and extinction series. The groups were close to identical during adaptation, yet the curves diverged during acquisition. The acquisition difference all but vanished during extinction. Analysis of variance of the adjusted GSRs revealed that the difference between the two groups was almost statistically significant, $F\ (1,\ 18) = 3.43$, $MS_e = .025$.

The mean GSR on those trials on which there was evidence of movement in the EMG records was compared with the mean GSR on the remaining trials. The GSR was slightly *larger* on the nonmovement trials. Twenty-four *S*s' mean GSRs were larger on those trials on which movement occurred and 24 *S*s' GSRs were larger on those trials without movement. The relationship between presence or absence of movement and whether the GSR avoidance criterion was met was examined for *S*s in the avoidance group. Deviations from expectation were very close to zero (chi-square $= .119$ for all 24 avoidance *S*s and chi-square $= .016$ for the 19 best-matched *S*s).

Discussion.—The results of this study tend to support Church's (1964) contention that the yoked control design may lead to spurious or exaggerated evidence of avoidance conditioning. When only pairs of *S*s who are very closely matched in both GSR responsivity and classical conditionability were used, differences favoring an avoidance group over its yoked controls tended to shrink. However, even when *S*s were very carefully matched, the acquisition difference previously found did not vanish completely. Since the manner in which it was necessary to conduct the experiment to obtain acceptable measures of responsivity and conditionability (especially the latter) probably reduced the possibility of achieving avoidance conditioning, support for Church's position does not mean that GSR avoidance conditioning is impossible. The *S*s could not avoid the shock until the third acquisition trial in this experiment. The inability to avoid on the first two acquisition trials may have interfered with the development of an avoidance response, especially since *S*s may demonstrate attenuation of the CR very early in classical

---

[2] The multiple correlation was .736. Responsivity and conditionability were essentially uncorrelated ($r = .08$).

GSR conditioning (Kimmel, 1966). For this reason, a better test of Church's argument could probably be accomplished using the eyelid response. The problem may be insoluble, however, since *S*s would have to be given enough classical conditioning trials to obtain an estimate of conditionability and more trials would be necessary for the eyelid reflex than for the GSR.

To the extent that avoidance conditioning of the GSR *was* marginally demonstrated in this study, no evidence of its dependence upon movement in the hand or arm containing the EMG and GSR electrodes could be identified. Since it has recently been shown (Kimmel & Davidov, in press) that simply pairing a CS and electric shock with a 5 sec. CS-UCS interval may not be adequate to produce EMG conditioning, the lack of an EMG-GSR relationship in this study is not surprising.

REFERENCES

CHURCH, R. M. Systematic effect of random error in the yoked control design. *Psychol. Bull.*, 1964, **62**, 122–131.

KIMMEL, H. D. Inhibition of the unconditioned response in classical conditioning. *Psychol. Rev.*, 1966, **73**, 232–240.

KIMMEL, H. D., & BAXTER, R. Avoidance conditioning of the GSR. *J. exp. Psychol.*, 1964, **68**, 482–485.

KIMMEL, H. D., & DAVIDOV, W. Classical GSR conditioning with concomitant EMG measurement. *J. exp. Psychol.*, 1967, in press.

KIMMEL, H. D., STERNTHAL, H. S., & STRUB, H. Two replications of avoidance conditioning of the GSR. *J. exp. Psychol.*, 1966, **72**, 151–152.

(Received May 31, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 1, 146–148

RESPONSE TRANSFER MEDIATED BY MEANINGFULLY SIMILAR AND ASSOCIATED STIMULI USING A SEPARATE-LISTS DESIGN

RICHARD A. KASSCHAU AND HOWARD R. POLLIO[1]

University of Tennessee

Ryan's (1960) comparison of verbal response transfer mediated by meaningfully similar and associated stimuli in a mixed-list design was repeated using separate lists for each type of mediational relationship. On most measures, meaningfully similar stimuli were found to be indistinguishable from associatively related stimuli in mediating response transfer. In contrast to Ryan, both types of interlist relations facilitated transfer more than unrelated control lists. These results were interpreted to mean that which set of stimulus relations Ss utilize in mediating transfer depends upon particular task constraints.

This study is a variation of an experiment reported by Ryan (1960) in which each *S* learned an original nine-word paired-associate list (OL) to a criterion of one correct anticipation. This was followed by a transfer list (TL) in which the same response words were paired, three each, with nine new stimulus words related to OL stimuli in the following ways: (*a*) Type A: primary word associations with minimum semantic similarity (boy-girl); (*b*) Type S: semantically similar but associatively unrelated words (health-vision); or (*c*) Type C: minimally related control words (slow-flaming). After learning TL, *S* was asked to relearn the original list (RL). Ryan concluded that although both Type A and Type S interlist relationships could facilitate transfer, word associations (Type A) were the major determiners of mediated transfer. Ryan's work has been extended through a change in method by Jenkins and Brown (1965), and in all important respects these new data support his earlier interpretation.

Both Ryan and Jenkins and Brown used a mixed-list design; a procedure that might have resulted in an unwanted interaction between Type A and Type S relations. That is, the mixed-list design might have lead *S* to expect an easily detected, high probability Type A relationship in all transfer pairs, thereby spuriously depressing *S*s' performance on Type S pairs.

We attempted to eliminate this possibility by presenting each *S* with a TL in which all

---

[1] This research was jointly supported by a University of Tennessee Nonservice Fellowship and Research Grant MH-08903 from the United States Public Health Service.

stimuli were related to OL stimuli in only one way; whether Type A, Type S, or Type C. By this procedure, the deleterious interactive effects of a mixed-list design were avoided, and a more unequivocal estimate of Type S transfer obtained.

Method.—The lists used were those reported by Ryan (1960). There were six groups of *S*s; two learned lists with Type A interlist stimulus relationships, two with Type S, and two with Type C. The two groups learning lists with a particular relationship were distinguished only in terms of whether the basic list, i.e., the list common to all groups, was learned as the original (Sequence I) or as the transfer list (Sequence II).

The procedure used was identical to that followed by Ryan with the following minor exceptions: (*a*) 60 undergraduates from introductory psychology courses were used, 10 per group. (*b*) The memory drum used did not have a shutter, hence to expose a stimulus-response pair the drum rotated at the end of the 2-sec. anticipation period. (*c*) The practice list consisted of five men's names paired with two-digit numbers.

TABLE 1

MEAN NO. CORRECT ANTICIPATIONS AND *SD*s

Transfer Relationship	Original List		Transfer List				Relearn Original List	
	All Trials		Trial 2		Trial 3		Trial 1	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Type A	69.5	25.5	5.5	2.3	5.9	2.3	8.2	0.9
Type S	66.3	20.7	4.0	1.9	4.8	2.1	7.4	1.4
Type C	62.4	25.1	2.6	1.5	4.3	2.1	6.4	0.9

Results and discussion.—Table 1 presents the mean number of correct anticipations and *SD*s for the OL, TL, and RL conditions. An analysis of variance indicated no significant differences in the total number of correct anticipations in OL as a function of either Sequence or Group effects. The amount of transfer obtained during TL learning was measured by the number of correct anticipations given on Trial 2 (the first anticipation trial) and on Trial 3. An analysis of variance of the Trial 2 data indicated that the three groups differed significantly, $F\ (2,\ 54) = 10.74, p < .001$. However, an analysis of the Trial 3 data indicated no significant differences among Groups, $F\ (2,\ 54) = 2.08, .10 < p < .20$. Duncan's multiple-range test applied to Trial 2 data indicated that Type A elicited more correct anticipations than either Type S ($p < .05$) or Type C ($p < .001$), and that Type S elicited more correct anticipations than Type C ($p < .05$). Analysis of the correct anticipations during Trial 1 in relearning revealed that the three groups differed significantly ($p < .001$), and that Sequence was also a significant effect ($p < .025$). Duncan's multiple-range test indicated that Type A elicited more correct anticipations than Type S ($p < .05$) and Type C ($p < .001$), while Type S elicited more correct anticipations than Type C ($p < .01$). In each of the above analyses, as well as those which follow, Type of transfer relationship (Groups) and Sequence were the primary factors. The

TABLE 2

MEAN TRIALS TO CRITERION, MEAN PERCENTAGE OF SAVINGS SCORES, AND *SD*s FOR ALL LEARNING CONDITIONS

Transfer Relationship	Learning Condition									
	Original		Transfer				Relearning			
	Trials to Criterion		Trials to Criterion		% Savings Scores[a]		Trials to Criterion		% Savings Scores[b]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Type A	14.4	6.5	7.6	4.2	43	28.3	2.2	2.4	84	11.0
Type S	13.1	6.8	8.0	6.4	41	22.8	3.2	2.0	75	12.0
Type C	11.1	4.7	11.6	9.8	−12	92.9	4.4	2.3	53	33.2

[a] $\frac{\mathrm{OL}-\mathrm{TL}}{\mathrm{OL}} \times 100$ = Percentage of savings.

[b] $\frac{\mathrm{OL}-\mathrm{RL}}{\mathrm{OL}} \times 100$ = Percentage of savings.

analyses involving Sequence were nonsignificant in all instances save two. Since no discernible pattern could be detected in these results, significant Sequence effects are reported but not discussed. Furthermore, no significant Sequence × Transfer Relationship interaction was ever found.

Table 2 shows the mean total number of trials to criterion and *SD*s for each list. Analysis of variance disclosed that the three groups did not differ significantly in OL performance, $F\ (2, 54) = 1.27,\ p > .20$. Although a trend is obvious in the TL condition, this difference did not achieve statistical significance, $F\ (2, 54) = 1.90,\ .10 < p < .20$. Analysis of the total trials to criterion in the RL condition, however, revealed a significant difference among Groups ($p < .025$). Duncan's multiple-range test indicated that Type A lists were relearned in significantly fewer trials than Type C lists ($p < .005$); no other difference was significant.

The data summarized in Columns 1, 3, and 7 of Table 2 were used to compute savings scores in transfer from OL to TL and from OL to RL. Considering first the savings scores for transfer from OL to TL (Columns 5 and 6 of Table 2), an analysis of variance indicated a significant difference as a function of the Transfer relationship ($p < .005$), as well as a significant Sequence effect ($p < .05$). Duncan's multiple-range test showed that Type A and Type S savings scores differed from Type C ($p < .01$), but did not differ significantly from each other. Further, an analysis of the savings scores from OL to RL (last two columns of Table 2) demonstrated significant differences in group performance ($p < .025$). Duncan's multiple-range test indicated that Type A and Type S did not differ significantly, whereas Type C elicited significantly lower savings scores than Type S ($p < .05$) and Type A ($p < .01$).

The results of this study are in agreement with earlier results in all respects except one: the magnitude of transfer obtained with Type S relations. In both previous studies associative factors were found to produce significantly better performance than semantic factors at all stages of transfer and relearning. Although savings measures were not explicitly reported, these earlier data also suggest significant differences between Type A and Type S pairs.

In terms of the present results only the early trials of the TL and RL conditions revealed significant differences favoring Type A pairs, with this difference much less striking in the RL than in the TL condition. By contrast, in terms of trials to criterion and savings score measures, both Type A and Type S relations produce significant positive transfer relative to Type C with no significant differences between Type A and Type S.

These data suggest that the relative efficacy of Type A and Type S relations in mediating transfer revealed in mixed-list designs may be somewhat artifactual: whatever greater effects accrue to Type A relations occur early in training and are easily compensated for by Type S relations after a small number of trials. This implies that *S* has at least two different types of relations available for mediating transfer: word associations and semantic similarity. Where a single task requires utilization of both word association and semantic mediators, *S*s show a preference for Type A relations and are delayed somewhat in using Type S relations. Where the task is such that only one type of relationship is available, there are initial differences in favor of word associations, with this advantage overcome almost immediately. Which set of relations *S*s utilize in mediating transfer would seem strongly to depend upon task constraints, although the present results indicate that both types of relations are capable of producing such effects.

REFERENCES

JENKINS, J. J., & BROWN, L. K. The use of interspersed test items in measuring mediated response transfer. *J. verbal Learn. verbal Behav.*, 1965, **4**, 425–429.

RYAN, J. J. Comparison of verbal response transfer mediated by meaningfully similar and associated stimuli. *J. exp. Psychol.*, 1960, **60**, 408–415.

(Received June 6, 1966)

Journal of Experimental Psychology

Vol. 74, Part 1 of No. 2 June 1967

DISCRIMINABILITY OF STIMULI VARYING IN PHYSICAL AND RETINAL ORIENTATION[1]

FRED ATTNEAVE AND RICHARD K. OLSON

University of Oregon

In 2 discrimination reaction-time experiments, adult human Ss responded faster to horizontal and vertical stimuli (lines or rectangles) than to stimuli tilted 45° right and left. When S viewed the stimuli with his head tilted 45°, so that physical and retinal orientation were in opposition, it was on the physical rather than the retinal horizontals and verticals that performance was superior. In another experiment head position was changed 45° after a period of learning. Ss required to give the same responses to the same physical orientations did much better on the transfer task than those required to give the same responses to the same retinal orientations. The latter were not significantly superior to a pure transposition group for whom the S-R relationships were shifted both physically and retinally.

The octopus discriminates easily between a horizontal and a vertical rectangle, but not between two oblique rectangles that also differ by 90° (Sutherland, 1957, 1958, 1960). Very similar results are obtained with preschool children (Rudel & Teuber, 1963); also with goldfish (Mackintosh & Sutherland, 1963). In the cat, however (Sutherland, 1963), no such difference is found, and Sutherland remarks in passing that the oblique stimuli seemed less confusable to the cats than to him. The most obvious explanation of these behavioral results is in terms of receptive field configurations, particularly since Young (1962) has observed that dendritic trees in the visual system of the octopus show proliferation most often in vertical and horizontal directions, and since Hubel and Wiesel (1959) have found that slope analyzers in the cat's visual cortex are about evenly distributed over all orientations. One is tempted to suppose, from the behavioral evidence, that man's primary visual system contains a preponderance of vertical and horizontal analyzers. However, no converging anatomical or physiological evidence exists at the human level at

---

[1] This research was supported by the Air Force Office of Scientific Research, Grant No. 973-66.

present, and the results reported here tend in fact to discourage this hypothesis in any simple form.

We asked the following questions: (*a*) In a discrimination reaction-time task, will human adults identify horizontal and vertical stimuli (lines or rectangles) more quickly than oblique ones? (Exp. I and III). It is well known that DRT increases as stimuli become more similar or confusable. (*b*) If so, is the effect dependent on the *physical* or the *retinal* orientation of the stimuli? (Exp. I and III). This was determined by running some *S*s with heads tilted 45°. (*c*) When people learn to identify slants, is it to the physical or to the retinal orientation that a response becomes attached? (Exp. II). This was determined by a transfer study.

Experiments I and II are conceptually discrete, but involve the same *S*s and overlapping data. Experiments I and III are conceptually similar, but differ in *S*s, method, and materials.

EXPERIMENT I

Method

Subjects.—The *S*s were 48 paid University of Oregon undergraduate volunteers, 30 females and 18 males between the ages of 18 and 25 yr. It was required that acuity of the right eye be 2/2 or better, as measured by a suitably reduced Snellen chart in the tachistoscope, and that response to a Lancaster-Regan figure show no suggestion of astigmatism.

Materials.—Stimuli were black lines, .21 in. long (.5° visual angle) × .02 in. wide, drawn on white detail paper which in turn was mounted on cardboard. Four orientations were used: horizontal, vertical, and two 45° diagonals. The center of each line was .84 in. (2°) from a central fixation point, which appeared only in the preexposure field as a black dot .1 in. in diameter on a white ground. Each slant was drawn, on separate cards, in eight different directions (above, upper right ... etc.) from the fixation point, making 32 cards in all. The reason for this variation was to avoid associating a given slant with unique retinal elements.

Practice stimuli were drawn in equally spaced positions around the edge of a circular white paper disk, which was fastened to a wall. The order was random, with each orientation appearing 11 times.

Apparatus.—A Gerbrands two-field tachistoscope was fitted with a rotatable head piece which held *S*'s head in an upright or 45° right position. The head piece blocked light from outside the tachistoscope, and allowed viewing with the right eye only. Both preexposure and stimulus fields were masked to a circular area 7 in. in diameter, 2 ft. from *S*. The (ground) luminance of both fields was 7½ mL. No other contours were visible inside the tachistoscope.

A hand switch held by *S* triggered the stimulus field for a 1/10-sec. exposure and also started a clock. The output from a microphone taped to *S*'s throat was led to a voice relay which shut off the clock.

Procedure.—The *S* was told that the purpose of the experiment was to study his reaction time under several different conditions. No other statement about the nature of the study was made.

After being tested for acuity and astigmatism, *S* was told that on each trial he would look into the tachistoscope, keeping his head fitted snugly into the head piece, and fixate the dot in the preexposure field. When he pressed the hand switch a short line would flash on the screen. The line would have one of four slants, each with a different name. The *S* was to say the name of the line as quickly as possible following its exposure. After this general orientation, he was shown the stimuli and told their respective names. He then practiced responding with these names to the stimuli arranged around the circle on the wall; incorrect responses were promptly corrected. Two cycles of the circular array constituted the preliminary training session.

The four names used were *Adam, Abner, Albert,* and *Andrew.* Assignment of these to line orientations was completely counterbalanced (in the Latin-square manner) over *S*s within each group (also within each of the small groups into which *S*s were divided in Exp. II). The idea of using manual rather than verbal responses was considered but rejected because almost any motor response is likely to have some relationship of compatibility or incompatibility with a line of a given slant.

The 48 *S*s were divided into two groups of 24. Group U viewed the stimuli with their heads upright; Group T viewed them with heads tilted 45° clockwise, in the preliminary session as well as in the experiment proper. Thus, for Group T, the *physically* oblique lines were *retinally* horizontal and vertical, and vice versa.

In the final instructions, *S* was told to press the button each time *E* gave a "ready" signal, and respond with the appropriate name as quickly as possible, without making mistakes. (Whenever an error was made, *S* was reprimanded, and the trial was repeated a little later in the series.) The *S* was allowed to sit back from the viewing position between trials and rest if he wished, but in practice he rarely did so. The surround of the head piece was a homogeneous screen the illumination of which was roughly matched to that of the tachistoscope fields. Between trials *E* changed stimuli, gave the "ready" signal, checked the correctness of the response, and recorded the reaction time: those operations made for an intertrial interval of about 20 sec. Two series of the 32 stimuli were presented for a total of 64 trials. Order was randomly permuted within each series.

Results

DRTs obtained under the different conditions are shown in Table 1. The terms *horizontal, vertical,* and *oblique* refer to physical rather than retinal orientation. Classification in terms of retinal orientation would have been equally legitimate; because of this basic ambiguity *t* tests on the results are rather more comprehensible than *F* tests would be. For Group U, reactions to horizontals and verticals are faster than to obliques by a small (72 msec.) but significant amount, t (23) $= 2.94$, $p < .01$. For Group T, the corresponding difference (re physical orientation) is 56 msec., which is just short of the .05 confidence level, t (23) $= 1.95$. The difference between these two differences (72 vs. 56 msec.) is quite insignificant, t (46) $= .43$. Note, however, that if we compare slants of equivalent *retinal* orientation, the sign of the 56-msec. difference for Group T becomes negative, and the difference between differences (72 vs. −56 msec.) becomes highly significant, t (46) $=$ 3.39, $p < .01$. These results clearly support the conclusion that discriminability is dependent on physical rather than retinal orientation, though the possibility that retinal orientation makes *some* difference is not disproved.

TABLE 1

MEAN DRTs IN MSEC., EXP. I

Stimuli (re physical axes)	Head Upright	Head Tilted
Obliques	1019	1129
Horizontals & Verticals	947	1073
Difference	72	56
Mean (all Slants)	983	1101

For unknown reasons (which might have to do merely with fatigue, annoyance, etc.) performance of Group T was generally poorer than that of Group U: the difference between overall means (983 and 1,101 msec.) is significant at the .05 level, t (46) $=$ 2.18.

EXPERIMENT II

We now ask: if an *S* learns to respond differentially to slants with his head in a given orientation, as in Exp. I, and his head position is then changed (from upright to 45° tilt, or vice versa), will he find it easier to give the old responses to stimuli with the same physical orientation, or to stimuli with the same retinal orientation? A transfer study of this sort, in which the entire procedure of Exp. I became the "original learning" portion, was conducted with the same *S*s.

Procedure.—Immediately after the 64 DRT trials described earlier, *S* was given new instructions dependent on the transfer group to which he was assigned. Group U was divided into Transfer Groups 1, 2, and 3 of eight *S*s each, likewise Group T into Transfer Groups 4, 5, and 6. The division was such as to preserve counterbalancing of names over slants within each transfer group.

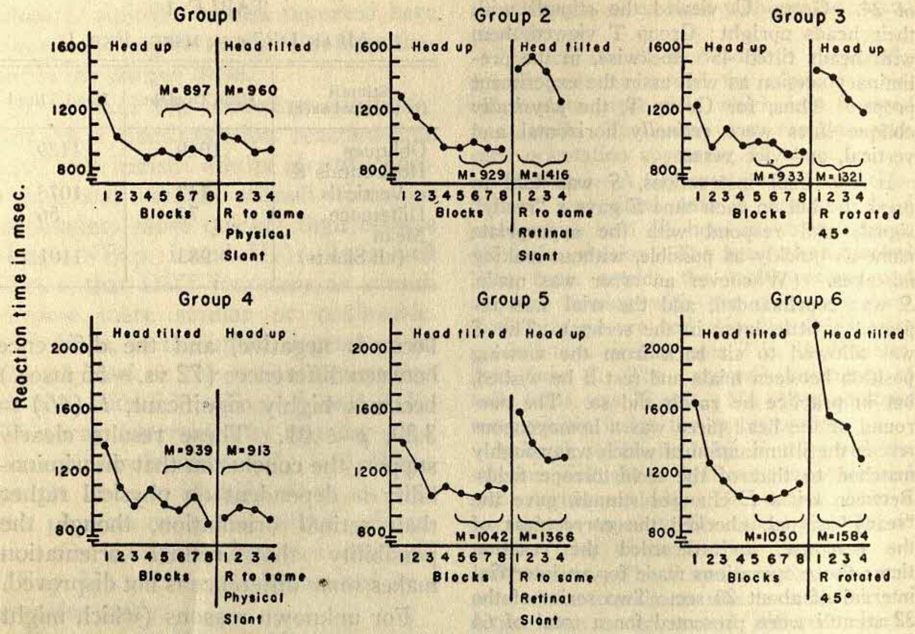


FIG. 1. Performance curves, pre- and posttransfer, for the six groups of Exp. II. (Each point is the mean of 8 consecutive trials. Numerical means are given for the 32 transfer trials and for the 32 trials immediately preceding transfer.)

In Groups 1 and 2, *S*'s head position was shifted from upright to tilt (45° clockwise); symmetrically Groups 4 and 5 were shifted from tilt to upright. In Groups 1 and 4 *S*s were told that they would use the same names they had been using for lines that were physically or objectively the same, "just as if you rotated your head while reading without rotating the book." The *S*s in Groups 2 and 5 were told to use the same names they had been using for lines that were in the same relation to their heads, "just as if you rotated your head while reading and rotated the book at the same time." For Groups 3 and 6 the transfer task was one of pure transposition: They maintained the same head position as in Exp. I, but the names were rotated 45° clockwise for the upright group (3) and 45° counterclockwise for the tilt group (6): "It's just as if you were reading and rotated the book without rotating your head." In all groups the new assignments of names to slants were demonstrated in detail, and *S*'s understanding of the changes was tested in part by having him name the first four lines on the practice disk. If his grasp of the transfer task seemed at all dubious, he was instructed further until it was clear that he understood the principle involved.

Each group was then run through a complete series of 32 randomly ordered trials (each stimulus card presented once) under the new conditions.

At the end of Exp. II, *S* was asked what he had come to think of as "up" in the new situation, in an attempt to determine whether any such shift in frame of reference had occurred.

Results

Figure 1 shows performance of each group under original (Exp. I) and altered (Exp. II) conditions. The points plotted are means for blocks of eight consecutive trials, with all slants combined. Transfer performance was compared with that on the second half only of the pretransfer trials; note that the learning curves in this region are very nearly asymptotic.

On the whole, these results present a fairly clear and coherent picture. When the same responses are attached to the same physical orientations, but to different retinal orientations (Groups 1 and 4), little or no performance decrement occurs. When the same responses are attached to the same retinal orientations but to different physical orientations (Groups 2 and 5), performance is markedly disrupted, much as it is in a pure transposition task (Groups 3 and 6), in which physical and retinal orientation are both changed.

An exhaustive evaluation of the results involves so many comparisons that we found it necessary to do several separate analyses of variance on the data. Since it seems unjustifiable to burden the reader with the details of these, we shall report only those outcomes for which some minimal interest or importance is evident.

The change in conditions produced no significant performance decrement in Groups 1 and 4; $F\ (1,\ 14) = .3$. The decrement is highly significant in Groups 2 and 5; $F\ (1,\ 14) = 27.4$, $p < .001$, and in Groups 3 and 6, $F\ (1,\ 14) = 40.6$, $p < .001$. In each of these analyses head position (or direction of change in head position, i.e., the difference between the two paired groups) constituted a second classification: in no case did this variable produce significant differences, either as a main effect or in interaction (all p's $> .1$). It may be pointed out that the latter tests are less powerful than the former, since they involve comparisons between rather than within *S*s.

A further analysis on within-*S* difference scores shows, as one might expect from the foregoing, that the decrement in Groups 2 and 5 is greater than in Groups 1 and 4: $F\ (1,\ 28) = 24.54$, $p < .001$.

Finally, we sought to determine whether the holding constant of retinal orientation provided an advantage over pure transposition. Groups 2 and 5 turn out not to differ significantly from Groups 3 and 6 with respect to within-*S* difference scores: $F\ (1,\ 28) = .315$. This was a two-way analysis of variance with the groups paired, somewhat arbitrarily, as in Fig. 1. If effects of absolute head position had caused differences to be increased in Group 2 and decreased in Group 5, this effect would have appeared in the interaction term. Actually the interaction was nonsignificant: $F\ (1,\ 28) = 2.45$, $p > .1$.

Answers to the question at the end of the experiment, "What did you think of as 'up'?" generally referred to the physical vertical. Only three *S*s indicated that they rotated their field orientations in the transfer task. These were all in Group 2, and their DRTs under the transfer condition were by far the lowest in that group. They showed a mean decrement of 131 msec., whereas the remaining five *S*s in Group II showed a mean decrement of 701 msec.

EXPERIMENT III

The differences in DRT attributable to slant in Exp. I were rather small (about $7\frac{1}{2}\%$), even with head upright. We wondered if the method (all four slants discriminated in the same series) might have been less than optimal for revealing slant effects. Another possibility considered was that all the lines used were pressing some asymptote of discriminability. We decided therefore to vary both method and materials in a new experiment. A two rather than four choice DRT paradigm was used (horizontal vs. vertical, or left oblique vs. right oblique); and 3:2 rectangles were used as well as lines (which may themselves be considered roughly 10:1 rectangles).

Method

Subjects.—Forty-eight new *S*s were employed, 22 men and 26 women. They were screened for acuity defects and astigmatism as in Exp. I.

Materials.—The 32 stimulus cards of Exp. I (q.v.) constituted half the materials. An additional 32 cards were constructed by exactly the same plan, but the lines were replaced with 9/40 in. × 6/40 in. rectangles cut from black paper. These were very slightly longer than the lines, which subtended half a degree of visual arc in the tachistoscope.

The 3:2 proportion was chosen with the aim of providing stimuli that would take longer to discriminate with respect to orientation than the 10:1 lines, but that could still be discriminated with practically no error. Pilot work indicated that rectangles only a little plumper than 3:2 would not meet the no-error criterion, presumably because of acuity limitations: i.e., *S* sometimes reported that he simply could not "see" the orientation of the rectangle.

Apparatus.—The tachistoscope used in Exp. I.

Procedure.—Half of the *S*s discriminated lines, the other half rectangles. Orthogonally, half viewed the stimuli with heads upright, the other half with heads tilted 45° clockwise. The *S* was given a total of 128 trials, in 8 blocks of 16. Within each subgroup, half the *S*s discriminated between horizontal and vertical orientations during odd blocks and between right oblique and left oblique during even blocks; for the other half this assignment was reversed. Within any given block of 16 trials, order of the 16 cards displaying horizontals and verticals, or on alternate blocks the 16 obliques, was randomly permuted.

TABLE 2

MEAN DRTs IN MSEC., EXP. III

Stimuli (re physical axes)		Head Upright		Head Tilted	
Obliques	Lines	591	628	592	645
	Rects.	664		698	
Horizontals & Verticals	Lines	549	584	573	625
	Rects.	619		676	
Differences	Lines	43	44	19	21
	Rects.	45		22	
Means (all Slants)	Lines	570	606	583	635
	Rects.	642		687	

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE, EXP. III

Source	*df*	*MS*	*F*
Between *S*	47		
A (Materials)	1	1853.44	6.462*
B (Head tilt)	1	200.25	<1
AB	1	64.88	<1
S w. Groups (Error, b.)	44	286.81	
Within *S*s	48		
C (Slant)	1	248.75	23.578***
AC	1	.06	<1
BC	1	33.03	3.131
ABC	1	.25	<1
C×*S* w. Groups (Error, w.)	44	10.55	

* $p < .05$.
*** $p < .001$.

The *S*'s alternative responses were always the two vowel sounds "o" and "e," which operated a voice key as before. Assignment of responses to stimuli was balanced within subgroups, though with a relational invariance that will be mentioned later.

The simplicity of the task was such that no extended pretraining was necessary; *S* was merely shown the stimuli in the tachistoscope and told the response to give to each. He was reminded of the correct responses at the beginning of each block.

Aspects of procedure not explicitly described in this section were as in Exp. I.

Results

The average DRTs obtained are shown in Table 2, and most but not all questions of significance are covered by an analysis of variance summarized in Table 3.

The rectangles take somewhat (about 15%) longer to discriminate than the lines; not as much longer as we had hoped in designing the experiment. Otherwise, results with the two types of materials are so closely parallel that they need not be considered separately.

The main effect for slant is significant at the .001 level. More important, Group T *S*s, considered separately, respond significantly faster to physical horizontals and verticals (i.e., retinal obliques) than to physical obliques (i.e., retinal horizontals and verticals): $t = 2.91$, $p < .01$. (In Exp. I this difference was just short of the .05 level. Differences attributable to slant are consistently smaller—either in milliseconds or as proportions of total DRT—in Exp. III than in Exp. I, but more significant because of reduced error.)

Consideration of the mean differences in Table 2 suggests that the advantage of the (physical) horizontals and verticals over the obliques was about twice as great for Group U as for Group T, i.e., that retinal orientation is of some importance.[2] However, the difference attributable to retinal slant (44 vs. 21 msec.) is of very dubious reliability: note F_{BC}, for which $.05 < p < .1$.

If we turn the classification about and ask how the differences between *retinal* orientations are affected by head position, we are in effect comparing a 44 msec. difference with a negative 21 msec. difference (cf. the similar comparison in Exp. I). This difference is highly significant: $F\ (1, 44) = 23.5$, $p < .001$.

Discussion

People identify the orientation of horizontal and vertical objects more quickly—though not a great deal more quickly, in absolute terms—than that of oblique objects. By far the major portion of this effect, if not all of it, must be ascribed to physical rather than retinal orientation. This is to say that the perceptual system makes allowance or correction for head position, on the basis of proprioceptive information, before the effect occurs. (In the present experiments the corrective information must have come from the labyrinths, muscles, or joints, or else from some purely cognitive level, since the eye was presented merely with a circular field regardless of head position.) If one undertakes to explain this effect in terms of receptive field orientations, he must suppose that the analyzers in question are located in a part of the visual system that is tied to the retina in a labile manner.

The evidence supporting a retinal-orientation effect is exceedingly weak ($p < .1$), and is confined almost entirely to Exp. III. For the sake of argument, however, let us entertain the possibility that retinal orientation did make a small difference. Orientation of low-level receptive fields might be responsible, but other plausible hypotheses also exist. Suppose that the phenomenal vertical and horizontal (we use the term *phenomenal* for want of a better one) are more easily discriminated than the phenomenal obliques. Suppose further that the phenomenal reference axes are normally the same as the physical, but that people have some limited ability to adjust or rotate them into correspondence with the axes of the (tilted) head, when there is some advantage of doing so. (At least three *S*s in Exp. II seem probably to have done this.) If an *S* with tilted head in Exp. III could have voluntarily tilted his reference axes into correspondence with the axes of his head on alternate blocks of trials, he would have had the advantage of always dealing with phenomenal verticals and horizontals. (Note that this argument does not apply to the four-choice situation of Exp. I, in which practically no evidence for the importance of retinal orientation was found.) The advantage for an *S* using this strategy would have been further enhanced by the fact that he could always have used the same response for the same phenomenal slant, since by an accident of counterbalancing the same response ("o" or "e") was common to the vertical and right oblique, for all Group T *S*s.

---

[2] Inspection of the differences in Table 2 tends to give one more subjective confidence than is warranted in the reliability of retinal orientation effects, because of the close agreement of results between materials. This agreement is far greater than we have any right to expect on the basis of within-subgroup variability: note that the *F*s for interaction with materials (AC and ABC, Table 3) are much less than one. We checked carefully the computation of these suspiciously low interaction terms and verified their accuracy. We then considered the possibility that the error term was artificially inflated by within-group counterbalancing (of block order and S-R pairing). Calculation of a new error term with the counterbalanced variables (which were perfectly confounded with each other) held constant made virtually no difference, however. The only conclusion remaining is that the minuteness of these interaction terms is due to chance.

The possible influence of retinal or even optical effects cannot be excluded, however. It has been found repeatedly that visual resolution is best for horizontal and vertical test objects (Higgins & Stultz, 1948, 1949; Leibowitz, 1953; Ogilvie & Taylor, 1958; Taylor, 1963). Informal observation indicates that this effect depends on retinal orientation: according to Higgins and Stultz (1948) "the perceptibility of the diagonal lines was improved, and the perceptibility of horizontal and vertical lines diminished when the observer merely tilted his head."

In the case of other studies that have shown superior discrimination in vertical and horizontal regions—e.g., the experiment of Leibowitz, Myers, and Grant (1955) on estimation of radial location—it is an open question whether physical or retinal orientation is the important variable. These studies, and certainly the one of Rudel and Teuber cited earlier, might profitably be repeated with a head-tilt condition.

Experiment II shows clearly enough that responses are normally attached to physical rather than retinal orientations, i.e., that the perceptual machinery makes correction for head position before associations are formed (Groups 4 and 5) or retrieved (Groups 1 and 2). These results are in general agreement with those of Rock (1956) and Rock and Heimer (1957), whose *S*s attempted, under several conditions of head tilt, to identify tachistoscopically presented complex objects that were either upright physically or upright on the retina. The physically (or "phenomenally") upright objects were clearly the easier.

Special attention is due the three *S*s in Group 2 who reported rotating the frame of reference with the head, and who performed dramatically better than their group mates under the transfer condition. Either these *S*s are exceptions to the rule that associations are attached to physical rather than retinal orientations, or else they were able voluntarily to adopt a retinal frame of reference when it became useful to do so. If the former hypothesis is true, why did not similar *S*s appear in Groups 1 and 4, where they would have shown very poor transfer? If the latter, why did not similar *S*s appear in Group 6, where rotation would have been equally advantageous? (In Group 5, voluntary adoption of a retinal frame of reference would have had to occur during the pretransfer trials, when *S* had no way of knowing that it could aid his later performance. In Group 3, *S* would have had to rotate to an arbitrary frame of reference, neither physical nor retinal.) About equally puzzling is the fact that the remaining five *S*s in Group 2 showed poorer transfer than any other group.

REFERENCES

HIGGINS, G. C., & STULTZ, K. Visual acuity as measured with various orientations of a parallel-line test object. *J. Opt. Soc. Amer.*, 1948, **38**, 756–758.

HIGGINS, G. C., & STULTZ, K. Variation of visual acuity with various test-object orientations and viewing conditions. *J. Opt. Soc. Amer.*, 1949, **40**, 135–137.

HUBEL, D. H., & WIESEL, T. N. Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.*, 1959, **148**, 574–591.

LEIBOWITZ, H. W. Some observations and theory on the variation of visual acuity with the orientation of the test object. *J. Opt. Soc. Amer.*, 1953, **43**, 902–905.

LEIBOWITZ, H. W., MYERS, N. A., & GRANT, D. A. Radial localization of a single stimulus as a function of luminance and duration of exposure. *J. Opt. Soc. Amer.*, 1955, **45**, 76–78.

MACKINTOSH, J., & SUTHERLAND, N. S. Visual discrimination by the goldfish: The orientation of rectangles. *Anim. Behav.*, 1963, **11**, 135–141.

OGILVIE, J. C., & TAYLOR, M. M. Effect of orientation on the visibility of fine wires. *J. Opt. Soc. Amer.*, 1958, **48**, 628–629.

Rock, I. The orientation of forms on the retina and in the environment. *Amer. J. Psychol.*, 1956, **69**, 513–528.

Rock, I., & Heimer, W. The effect of retinal and phenomenal orientation on the perception of form. *Amer. J. Psychol.*, 1957, **70**, 493–511.

Rudel, R. G., & Teuber, H.-L. Discrimination of direction of line in children. *J. comp. physiol. Psychol.*, 1963, **56**, 892–898.

Sutherland, N. S. Visual discrimination of orientation by *Octopus*. *Brit. J. Psychol.*, 1957, **48**, 55–71.

Sutherland, N. S. Visual discrimination of the orientation of rectangles by *Octopus Vulgaris Lamarck*. *J. comp. physiol. Psychol.*, 1958, **51**, 452–458.

Sutherland, N. S. Visual discrimination of orientation by *Octopus*: Mirror images. *Brit. J. Psychol.*, 1960, **51**, 9–18.

Sutherland, N. S. Cat's ability to discriminate oblique rectangles. *Science*, 1963, **139**, 209–210.

Taylor, M. M. Visual discrimination and orientation. *J. Opt. Soc. Amer.*, 1963, **53**, 763–765.

Young, J. Z. Memory mechanisms of the brain. *J. ment. Sci.*, 1962, **108**, 120–132.

(Received July 1, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 158–160

SLEEP CYCLING WITHIN TWENTY-FOUR HOUR PERIODS[1]

WILSE B. WEBB AND HARMAN W. AGNEW, JR.
University of Florida

Ss were studied while sleeping in the laboratory during the afternoon hours. The sleep of these Ss neither resembled that found during the last 2 hr. of nocturnal sleep nor that found in the first 2 hr. The data indicated that the amount of Stage 4 and 1 REM found during an interjected sleep event is a function of its proximity to the onset or the termination of the regular sleep period.

In studies of the electroencephalographic (EEG) parameters associated with experimentally uninterrupted nocturnal sleep, a differential distribution of Stage 4 and 1 rapid eye movement (1 REM) was noted (Williams, Agnew, & Webb, 1964, 1966). These normative studies demonstrated that the major portion of Stage 4 sleep occurred during the first one-third of the night. Stage 1 REM occurred in an increasing amount toward the last third of the night. Figure 1 presents an analysis of the distribution of these two stages for 3 nights of sleep on 40 *S*s. This differential distribution of stages has been called the "Intrasleep Cycle" (Webb, Agnew, & Sternthal, 1966).

In a recent study the proportion of Stages 4 and 1 REM found in the early morning after a night of sleep was reported (Webb et al., 1966). These *S*s began sleeping in the laboratory at 9 AM after having awakened at 7 AM from nocturnal sleep. During the 2-hr. sleep period they showed a very high absolute and proportional amount of Stage 1 REM and a very low amount of Stage 4 sleep. Instead of recycling as at the beginning of nocturnal sleep, the sleep pattern occurring shortly after awakening more closely resembled the last 2 hr. of a night's sleep. These data indicated that the recycling of sleep does not necessarily occur with sleep onset. Rather, it appears that the amounts of Stages 4 and 1 REM which occur are a function of the proximity of the interjected sleep period to the major sleep period, or they are a function of the amount of deprivation following the sleep period. This study extends this early morning sleep study and is concerned with the distribution of Stages 4 and 1 REM as they occur in the afternoon.

METHOD

Three groups, each consisting of 12 young adult males (17–21 yr. old), were studied during the afternoon hours. On the first afternoon, for purposes of adaptations, *S*s were wired for the EEG and retired for a 3-hr. sleep period during which no record was obtained. This afternoon of sleep was

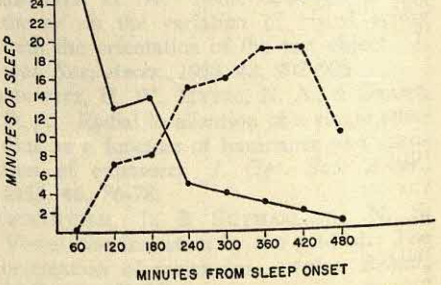


FIG. 1. Differential distribution of Stages 4 and 1 REM noted across the entire night of sleep for 3 nights of sleep on 40 *S*s.

---

[1] Supported by Air Force Office of Scientific Research Grant AF-AFOSR-395-63.

TABLE 1

PERCENTAGE OF TIME SPENT IN EACH SLEEP STAGE DURING AFTERNOON SLEEP

Stage	Last 2 Hr. of Night's Sleep	9:00 AM[a]	12:30 PM	2:00 PM	4:00 PM	First 2 Hr. of Night's Sleep[b]
0	1	4	11	8	17	1
"1"	7	12	14	9	11	6
1 REM	39	32	12	15	8	3
2	47	48	44	42	37	36
3	3	4	4	7	6	11
4	3	0	15	18	20	43

[a] Data from Webb et al., 1966.
[b] Data from Williams et al., 1964.

provided in an effort to wash out any "first night effect" (Agnew, Webb, & Williams, 1966). On the second afternoon 17 electrodes were applied to the scalp using the montage and procedures described by Williams et al. (1966). Depending upon the group, *S* retired to bed and was given the first opportunity to sleep at either 12:30 PM, 2:00 PM, or 4:00 PM. He was told to sleep as long as possible but that in any case the recording period would last for 3 hr. Subsequently an eight-channel EEG and a two-channel EOG were recorded for 3 hr.

The resulting records were scored for five stages of sleep using a modified version of the Dement-Kleitman system. These criteria are fully explained elsewhere (Williams et al., 1964) but a summary is as follows: Stage 0 contains alpha wave occipital wave activity and is indicative of the waking state. Stage 1 is the low voltage fast stage of sleep. When accompanied by a recording of rapid eye movements this stage is scored as 1 REM. Stage 2 is marked by spindles in the EEG record. Stage 3 is a transitional stage containing some delta waves. Stage 4 is the high voltage slow wave stage of sleep with over 30 sec. of delta waves.

The statistical analyses were performed on only the first 120 min. of record after the onset of the first minute of Stage 1 sleep.

RESULTS

Table 1 displays the percentage of time each group of *S*s spent in the various stages of sleep in the afternoon. When these data are compared with the last 2 hr. of sleep in the morning and the first 2 hr. of sleep at night the differences in Stage 4 and 1 REM are quite apparent.

DISCUSSION

These results indicate that when the normal sleep-wakefulness pattern is maintained the distribution of Stage 4 and 1 REM in an interjected sleep event is a function of its proximity to the onset or the termination of the regular sleep period. If the interjected period approaches the onset of the regular period, the Stage 4 and 1 REM distribution will approximate the first portion of the regular sleep cycle. If, on the other hand, the interjected period is proximal to the termination of the regular period there will be a dominant pattern of 1 REM with limited Stage 4 as is shown for morning sleep in Table 1.

What can be said about cycling of sleep in cases other than those in which the regular sleep-wake pattern is maintained? So far as can be ascertained, no study has been made of the effect of "interjected sleep periods" on the regular sleep cycle as we have used this term. A few studies have noted the initial onset characteristics of the sleep cycle after partial, differential, and total deprivation. From a partial deprivation study in which *S*s were permitted only 3 hr. of sleep per night for 7 nights Webb and Agnew (1965) found that *S*s tended to cycle their sleep onset in a manner similar to their normal cycling patterns (dominant Stage 4 and limited Stage 1 REM) with the exception of a tendency to increase the amount of Stage 4 at the expense of Stage 3. Similar results have been noted in studies by Dement and Greenberg (1966) and Sampson (1965). On the "recovery" night of this experiment Stage 4 continued to dominate the early part of the sleep period. The studies of total deprivation, although they have not been oriented toward sleep cycling, have shown accord in noting the early onset and dominance of Stage 4 in the sleep period immediately following deprivation (Berger & Oswald, 1962). Finally, deprivation experiments (deprivation of Stage 4 or 1 REM) have in general shown that subsequent sleep is characterized by a large amount and an early onset of the stage which has been deprived (Agnew et al., 1964; Dement, 1960).

REFERENCES

AGNEW, H. W., JR., & WEBB, W. B. The first night effect: An EEG study. *Psychophysiol.*, 1966, **2**, 263–266.

AGNEW, H. W., JR., WEBB, W. B., & WILLIAMS, R. L. The effects of stage four sleep deprivation. *Electroencph. Clin. Neurophysiol.*, 1964, **17**, 68–70.

BERGER, R. J., & OSWALD, I. Effects of sleep deprivation on behaviour, subsequent sleep, and dreaming. *J. ment. Sci.*, 1962, **108**, 457–465.

DEMENT, W. C. The effect of dream deprivation. *Science*, 1960, **131**, 1705–1707.

DEMENT, W. C., & GREENBERG, S. Changes in total amount of stage four sleep as a function of partial sleep deprivation. *Electroenceph. Clin. Neurophysiol.*, 1966, **20**, 523–526.

SAMPSON, H. Deprivation of dreaming sleep by two methods. *Arch. Gen. Psychiat.*, 1965, **13**, 79–86.

WEBB, W. B., AGNEW, H. W., JR. Sleep: Effects of a restricted regime. *Science*, 1965, **150**, 1745–1747.

WEBB, W. B., AGNEW, H. W., JR., & STERNTHAL, H. Sleep during the early morning. *Psychon. Sci.*, 1966, **6**, 277–278.

WILLIAMS, R. L., AGNEW, H. W., JR., & WEBB, W. B. Sleep patterns in young adults: An EEG study. *Electroenceph. Clin. Neurophysiol.*, 1964, **17**, 376–381.

WILLIAMS, R. L., AGNEW, H. W., JR., & WEBB, W. B. Sleep patterns in the young adult female: An EEG study. *Electroenceph. Clin. Neurophysiol.*, 1966, **20**, 264–266.

(Early publication received
November 21, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 161–166

SEQUENTIAL VERSUS NONSEQUENTIAL VARIABLES IN PARTIAL DELAY OF REWARD[1]

E. J. CAPALDI

University of Texas

The experiment employing 44 rats in the straight alley was a 2 × 2 factorial containing 2 levels of percentage of delayed reward trials (40% vs. 60%) and 2 levels of D length (1 vs. 1 and 3), D length being defined as the number of delayed reward trials occurring in succession without interruption by an immediately rewarded trial. Running speed in acquisition was faster for 40% delay than for 60%, especially in the goal area, and resistance to extinction was greater for long D length (1 and 3) than for short (1). D length did not affect acquisition performance and percentage of delay did not affect extinction performance.

A schedule of partial delay of reward contains an intermixture of delayed reward (D) and immediately rewarded (I) trials and may be said to contain two general classes of variables, nonsequential and sequential. Nonsequential variables are independent of the specific sequence of occurrence of D and I trials. Examples are percentage and number of D trials. Two examples of variables dependent upon the sequence of D and I trials are D length, defined as the number of D trials occurring in succession without interruption by an I trial, and the number of different D lengths. The present investigation attempted to determine the relative influence of sequential and nonsequential variables on acquisition and extinction under partial delay. This was accomplished by employing a 2 × 2 factorial design containing two levels of percentage of delay (40% vs. 60%) and two levels of D length (1 vs. 1 and 3). Thus one of the 40% groups and one of the 60% groups received D lengths no greater than 1 while the remaining 40% and 60% groups received D lengths of 1 and of 3.

Some current approaches, e.g., the dissonance hypothesis (Lawrence & Festinger, 1962), are clearly nonsequential in character. Thus, according to the dissonance hypothesis, resistance to extinction is determined by number of D trials regardless of the sequence of D and I trials. Accordingly, the dissonance hypothesis would expect resistance to extinction to vary as a function of percentage of D trials and to be independent of D length. Other hypotheses (e.g., Capaldi, 1966) are sequential in character and would expect resistance to extinction to be affected by D length and to be independent of percentage of D trials.

Another purpose of the present investigation was to determine if delay of reward and partial reward could be understood in terms of the same general principles, an issue which has received increasing attention in recent years. Recently, for example, Marx, McCoy, and Tombaugh (1965) employing delay of reward obtained results which in their view suggest

---

[1] This investigation was supported in part by Research Grant HD 00949-03 from the National Institute of Child Health and Human Development. Gratitude is expressed to Hugh Poynor who ran *S*s and assisted with the statistical analyses.

"... the delay-extinction phenomenon cannot simply be attributed to a partial-reinforcement effect and may well require a different kind of theoretical explanation [p. 334]." Recent partial reward investigations (Capaldi, 1964; Gonzalez & Bitterman, 1964) have shown that long runs of nonrewarded trials produce greater resistance to extinction than shorter runs of nonrewarded trials when percentage of reward is identical. If long runs of D trials are also shown to produce greater resistance to extinction, then to that extent a similarity between delay of reward and partial reward will have been established.

METHOD

Subjects.—The *S*s, 120 days old on arrival at the laboratory, were 44 male rats purchased from the Holtzman Company, Madison, Wisconsin.

Apparatus.—The runway was 82 in. long, 4 in. wide, with 9-in. high sides, covered with ½-in. hardware cloth. An 8-in. long floor treadle, suspended over a microswitch, constituted the initial portion of the alley, and when the treadle was depressed by the rat, whose front paws were always placed at its extreme forward edge, the first clock (.01 sec.) was started. Interrupting a photobeam 2 in. beyond the treadle stopped Clock 1 (start speed) and started Clock 2. Interrupting a photobeam 52 in. beyond the first photobeam stopped Clock 2 (run speed) and started Clock 3. Interrupting a photobeam 15.5 in. beyond the second photobeam and 2.5 in. in front of a brass 2 × 4 × 1.5 in. food cup covered by a tightly fitting sliding, metal lid stopped Clock 3 (goal speed) and, depending on the setting of a series of Hunter timers, activated an electric motor which opened the lid covering the food cup, either immediately or 30 sec. later. The brass food cup contained a wooden insert with a small hole drilled in its center which accommodated the six .045-gm. Noyes pellets employed as reward on all trials whether immediately rewarded or delayed. The *S* was removed after consuming the pellets and placed in a neutral box for 15–20 sec. to await the next trial. Following the last trial of the day, *S* was returned to the home cage where it was fed the daily ration.

Preliminary training.—On arrival at the laboratory *S*s were placed in individual cages and were allowed complete access to food and water. On Day 4 the deprivation schedule maintained throughout the course of the experiment was begun, *S* receiving 13–15 gm. of Wayne Lab Blox pellets daily. On Days 9 and 10 *S*s were placed on a large table in groups of 10–12. The *S*s remained on the table, freely supplied with .045-gm. Noyes pellets, for 30 min. (Day 9) or 45 min. (Day 10) and were handled from time to time. On Day 11 each *S* explored the unbaited but otherwise fully operational runway (clicking relays, etc.) for 6 min. Beginning on Day 11 each *S* received the daily ration about 10 min. later than the estimated time of running for that *S* in experimental training. On Days 12 and 13 each *S* was placed in the goal box twice daily for 2 min., the placements being separated by a 1-min. interval spent in a neutral box. On these placements the lid covering the food cup was in the "open" position and the food cup was liberally supplied with .045-gm. pellets. On the two placements occurring on Days 14 and 15 the lid covering the food cup was in the closed position and opened automatically, making a characteristic, slight noise, as *E*'s hand and *S*'s body swept past the photobeam. On these days the food cup contained six pellets and *S* was removed after eating them.

Acquisition training.—There were 20 days of acquisition training at five trials each day. Each group of 11 *S*s each was trained under four different five-trial schedules of partial delay, each schedule occurring once in a 4-day block on a regular basis, i.e., Schedule A occurred on Days 1, 5, 9, 13, 17; Schedule B on Days 2, 6, 10, 14, 18; Schedule C on Days 3, 7, 11, 15, 19; Schedule D on Days 4, 8, 12, 16, 20. Each of the five-trial schedules is shown in Table 1. I represents immediate reward; D, a 30-sec. delay-of-reward trial.

For the groups labeled 60, 60% of the trials were delayed-reward trials, and for the groups labeled 40, only 40% of the trials were delayed. The symbols S and L refer to short and long, respectively, and represent D length, or the number of delayed reward trials occurring consecutively within a day's trials without interruption by an immediately rewarded trial. The column headed "DL" lists the D length occurring on that day's trials and (in parentheses) the number of times the particular D length occurred on that day's trials.

TABLE 1

SCHEDULE OF REWARD RECEIVED BY EACH OF THE FOUR GROUPS ON EACH OF THE 20 DAYS OF ACQUISITION TRAINING

Schedule	Group							
	60-S		60-L		40-S		40-L	
	Trial	DL	Trial	DL	Trial	DL	Trial	DL
A	IIDDD	—	DDDII	3(1)	IIIDD	—	DDDII	3(1)
B	DIDID	1(2)	IDDDI	3(1)	DIDII	1(2)	IDDDI	3(1)
C	DIIDD	1(1)	DIIDD	1(1)	DIIID	1(1)	DIIII	1(1)
D	IDIDD	1(1)	IDIDD	1(1)	IDIID	1(1)	IDIII	1(1)

In examining Table 1 it should be understood that a D trial or several consecutive D trials not followed by an I trial *on a particular day* are not to be counted as D lengths. This is because the stimulus produced by the D trial (or several D trials) is not conditioned to the instrumental response on that day's trials. While it is true that the delay stimulus in question will be conditioned to the instrumental response by an I trial occurring on the trials of the subsequent day, the effect of such conditioning can be expected to be minimal as Capaldi and Poyner (1966) have shown.

Close attention to the 2 × 2 factorial design will show that (*a*) in a 4-day block each group received the same number of transitions from delayed trials to immediately rewarded trials (four); (*b*) one of the 40% groups received D lengths of one and three, the other, D lengths of only one, the same being true for the 60% groups; (*c*) for Group 40-L percentage of delayed trials was 40% over a 4-day block; and (*d*) for all groups the initial daily trial was equally often D or I.

Extinction.—In extinction all conditions were the same as in acquisition, five trials each day, etc., except that *S* was confined to the goal box for 30 sec. on each trial and the lid covering the six pellets contained in the food cup remained closed throughout the 30-sec. confinement period. In both acquisition and extinction *S* was allowed 60 sec. to traverse each section of the alley. If 60 sec. elapsed in one section of the alley, *S* was guided to the next section, and from the goal section *S* was guided to the goal box.

RESULTS AND DISCUSSION

Acquisition.—All times were converted to inches per second. The largest differences occurred in the goal section, although differences in the run section were substantial. Goal speeds in 2-day blocks in acquisition and on each day of extinction are shown in Fig. 1; the left panel showing the effects of the length (L) variable, the right panel, the effects of the percentage (P) variable. Figure 1 shows that acquisition differences due to L were small with extinction differences being large, whereas P produced the opposite: relatively large acquisition differences and relatively small extinction differences. A 2 × 2 analysis of variance ($df = 1, 40$ in all cases) was

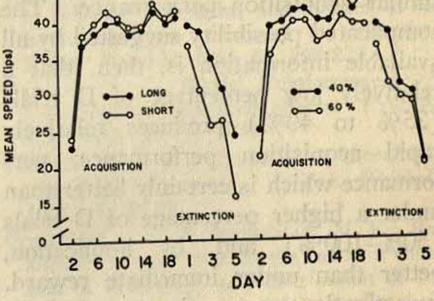


FIG. 1. Running speed in the goal area, in terms of inches per second on the 20 days of acquisition training (2-day blocks) and on each of the 5 days of extinction, as a function of D length (left panel) and percentage of delayed reward trials (right panel).

performed over the speeds on the last 10 days of acquisition in each alley section. Differences due to L were not significant, $F < 1$ in all sections, nor were differences due to the P × L interaction, $F = 1.20$, $F < 1$, and $F = 1.78$, for start, run, and goal, respectively. The present investigation is apparently the first to deal with the effects of D length on acquisition. Differences due to P were small in the start section, $F = 1.49$, approached significance in the run section, $F = 3.26$, $p < .10$, and definitely reached significance in the goal section. $F = 5.96$, $p < .025$. This finding supports Wike and McNamara (1957) who reported that 25% D trials produced faster running in acquisition than either 50% or 75% D trials. Wike and McNamara reported log running times along the total length of the alley. The present results and those of Wike and McNamara ostensibly support the generalization that the lower the percentage of D trials, the better the acquisition performance. However, numerous investigations have shown (Capaldi & Poynor, 1966; Logan, Beier, & Kincaid, 1956; Scott & Wike, 1956; Wike & McNamara, 1957) that 50% D trials and immediate reward produce similar acquisition performance. The complicated possibility suggested by all available information is, then, that a relatively low percentage of D trials (25% to 40%) produces relatively rapid acquisition performance, performance which is certainly better than under a higher percentage of D trials (50%–100%) and, by implication, better than under immediate reward. Investigations comparing a relatively low percentage of D trials (25%–40%) with immediate reward would appear to be in order.

Extinction.—A 2 × 2 analysis of variance ($df = 1, 40$ in all cases) was performed on the speeds over the 5 days of extinction in each alley section. Differences due to P were not significant, $F < 1$, $F < 1$, and $F = 3.69$ for start, run, and goal, respectively; nor were differences due to the P × L interaction, $F = 2.87$, 3.07, and 3.05, for start, run, and goal, respectively. Differences due to L, while not significant in the start section, $F < 1$, were highly significant both in the run, $F = 12.71$, $p < .001$, and in the goal section, $F = 19.39$, $p < .001$.

DISCUSSION

An extension of a hypothesis originally applied to partial reward (Capaldi, 1964) will serve to explain the present extinction results. Assume first that delayed reward produces a characteristic stimulus which is available for conditioning to the instrumental response on the subsequent trial (Capaldi & Poynor, 1966). Assume further that this characteristic stimulus is modified by successive D trials such that the stimulus consequent of a single D trial ($S^{D}1$) is different from that of two successive D trials ($S^{D}2$), and so on, with this modification occurring according to a negative growth function. It has already been suggested (Capaldi, 1964) that the stimulus produced by nonreinforcement (S^{N}) is modified by successive nonre-

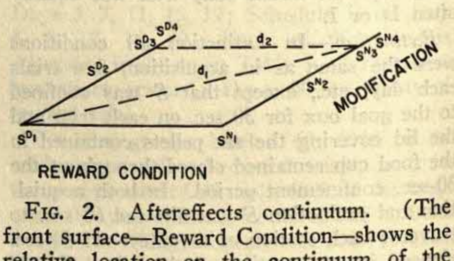


FIG. 2. Aftereffects continuum. (The front surface—Reward Condition—shows the relative location on the continuum of the aftereffect of delayed reward—S^{D}—and of nonreward—S^{N}. The portion of the continuum running from front to back—Modification—shows the modification undergone by the aftereffect as a function of successive delayed trials or successive nonrewarded trials.)

inforced (N) trials. The manner in which these propositions may be employed to explain the present results is illustrated in Fig. 2.

Figure 2 represents a two-dimensional surface which shows reward conditions (delayed reward or nonreward) along the front portion of the continuum and stimulus modification as a function of successive D or N trials running from the front to the back portion of the continuum. Figure 2 shows that habit strength generated at S^D1 generalizes to S^N3 along the path labeled d_1 with habit generated at S^D3 generalizing to S^N3 along the path labeled d_2. Since the distance d_2 is less than the distance d_1, it follows that S^N3 will receive more generalized habit strength from S^D3 than from S^D1. Actually, all higher values of S^N occurring in extinction, e.g., S^N10, S^N15, etc., will receive more generalized habit strength from S^D3 than from S^D1. Now, if both S^D1 *and* S^D3 possess habit strength, S^D3 and other higher values of S^N will receive more generalized habit strength than from S^D1 alone or from S^D3 alone because the habit strengths from the two generalized tendencies will summate (Capaldi, 1964). It follows from this analysis that the two long groups (40% + 60%) for which S^D1 and S^D3 were conditioned to the instrumental response should have been more resistant to extinction than the two short groups for which only S^D1 was conditioned. The expected result was obtained, L being highly significant in the run and goal sections.[2]

According to the present hypothesis percentage of D trials should have no effect on resistance to extinction, and this seems to have been the case. First, the somewhat substantial but nevertheless nonsignificant extinction effect of the P variable in the goal section would seem to have been occasioned by the faster running in acquisition produced by 40% delay relative to 60% delay. Note that in the run section where acquisition differences were smaller, the P variable had no effect on extinction, $F < 1$. Thus, it would appear that the results of Wike and McNamara (1957), which indicated generally that resistance to extinction increased as percentage of D trials increased, were in fact due to D length since under a quasirandom assignment of D and I trials, as occurred in the Wike and McNamara investigation, D length tends to increase as percentage of D trials increases. The present extinction results for partial delay are entirely consistent with results reported for partial reward (Capaldi, 1964; Gonzalez & Bitterman, 1964) which show that while percentage of reward fails to affect resistance to extinction, N length does, the N length variable corresponding to the D length variable of the present investigation except that delayed trials are replaced by nonrewarding trials.

REFERENCES

CAPALDI, E. J. Effect of N-length, number of different N-lengths, and number of reinforcements on resistance to extinction. *J. exp. Psychol.*, 1964, **68**, 230–239.

CAPALDI, E. J. Partial Reinforcement: A hypothesis of sequential effects. *Psychol. Rev.*, 1966, **73**, 459–477.

CAPALDI, E. J., & POYNOR, H. Aftereffects and delay of reward. *J. exp. Psychol.*, 1966, **71**, 80–88.

GONZALEZ, R. C., & BITTERMAN, M. E. Resistance to extinction in the rat as a function of percentage and distribution of reinforcement. *J. comp. physiol. Psychol.*, 1964, **58**, 258–263.

LAWRENCE, D. H., & FESTINGER, L. *Deterrents and reinforcement.* Stanford: Stanford Univer. Press, 1962.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and var-

---

[2] This deduction assumes equality or near equality of habit strength of the various aftereffects conditioned in this investigation (cf. Capaldi, 1964). That is, the habit strength available at S^D1 and S^D3 in the long groups should be equal or nearly equal to the habit strength available at S^D1 in the short groups. Such equality of habit strength was insured here, as in a previous investigation (Capaldi, 1964), by employing considerable acquisition training prior to extinction. Thus, it may be assumed that all S-R tendencies were at a near asymptotic strength and therefore possessed equal or near equal strength.

ied reinforcement. *J. exp. Psychol.,* 1956, **49,** 260–266.

Marx, M. H., McCoy, D. F., & Tombaugh, J. W. Resistance to extinction as a function of constant delay of reinforcement. *Psychon. Sci.,* 1965, **2,** 333–334.

Scott, E. D., & Wike, E. L. The effect of partially delayed reinforcement and trial-distribution on the extinction of an instrumental response. *Amer. J. Psychol.,* 1956, **69,** 264–268.

Wike, E. L., & McNamara, H. J. The effects of percentage of partially delayed reinforcement on the acquisition and extinction of an instrumental response. *J. comp. physiol. Psychol.,* 1957, **50,** 348–351.

(Received June 7, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 167–171

PROACTIVE AND RETROACTIVE INHIBITION IN THE RETENTION OF A T-MAZE HABIT IN RATS[1]

ROBERT G. CROWDER
Yale University

128 rats were tested for retention of a T-maze habit immediately, 1 hr., 1 day, or 1 wk. following massed acquisition. Control Ss learned only 1 position habit whereas experimental Ss learned 2 and later relearned either the 1st (RI paradigm) or 2nd (PI paradigm). The results showed no effects of prior learning upon retention for the 2nd habit at any interval. There were, however, severe retention losses attributable to interpolated learning and these losses tended to decrease with the length of the retention interval. While consistent with other studies of forgetting in lower animals, these findings were not comparable to the usual results of human verbal learning experiments.

As several writers have observed, the interference theory of human verbal learning and memory has drawn heavily upon concepts developed originally in the animal-conditioning laboratory (Gleitman & Jung, 1963; Kehoe, 1963; Rickard, 1965). Although it is not crucial that when the operations of verbal learning research are redirected to experiments on lower animals similar results obtain, the presence of gross discrepancies should at least be a source of concern and further research. Attempts to demonstrate the principal predictions of interference theory with lower animals have, in the past, yielded generally mixed (Kehoe, 1963) or negative results (Gleitman & Steinman, 1963; Gleitman, Steinman, & Bernheim, 1965; Rickard, 1965). In fact, proactive inhibition (PI) has only once been shown unequivocally (Gleitman & Jung, 1963).

The absolute number of theoretically based studies of retention in animals is, however, lamentably small. What reports there are range over several different species, different learning tasks, and different choices of retention intervals. Before undertaking any systematic conclusions about the contradictions among these studies and the contradictions between them and the verbal learning literature, it seemed important to obtain more data of a parametric nature. The present experiment was thus designed to provide comprehensive information on the forgetting of T-maze position habits by rats as a function of prior learning, interpolated learning, and retention interval. The study incorporated more adequate control groups, more retention intervals, and more *S*s than had previously been available in the literature.

METHOD

Design.—Sixteen groups of eight *S*s each were defined by factorial combinations of two paradigms (RI, PI), two interference conditions (single habit for controls, two habits for experimentals), and four retention intervals (immediate, 1 hr., 1 day, and 1 wk.). Experimental *S*s all received identical treatment during acquisition, but the first habit was later tested for RI groups and the second for PI groups. The distinction between RI and PI control *S*s was based on procedures which rendered them comparable to RI and PI experimentals, respectively, in

---

[1] This investigation was supported in part by research funds from Yale University and in part by Grant GB 4066 from the National Science Foundation. Portions of this material were presented at Eastern Psychological Association meetings on April 16, 1966 in New York City.

terms of quantity of reward and detention history during acquisition (see below for details). Each of the 16 groups was balanced, also, with respect to the direction retained and later tested (left or right), although these subgroups were not treated separately in the analysis of results.

Subjects.—The *S*s were 128 naive male albino Sprague-Dawley rats (Holtzman Farms), between 90 and 120 days old at the time of their service in the experiment. During their first week in the laboratory, *S*s' weights were reduced to, and subsequently maintained at, 85% of their ad lib feeding weights. The restricted ration of approximately 10-gm. Purina chow was given to all animals after each daily training session. Water was available in the home cages at all times. Before serving in the experiment, *S* was given a week's daily handling and one session's practice in accepting reward pellets (.045 gm. Noyes) from the food cup used in the T maze. Eleven *S*s of the original sample were replaced due either to failure to reach criterion on OL in 1 hr., sickness, or *E*'s error.

Apparatus.—The apparatus was a 3 in. (wide) × 4 in. (high) T maze with a 24 in. stem and 24 in. arms. The interior of the maze was lined with dull sheet metal, except for Plexiglas hinged lids covering all segments. A 12-in. starting chamber was separated from the rest of the stem by a guillotine door as were goal boxes by similar doors 12 in. from the end of each arm. Starting and choice latencies were recorded from clocks associated with photoelectric cells at the starting-box door and the goal-box doors, respectively. All light came from a single overhead 60-w. bulb.

Procedure.—Acquisition training for all *S*s was accomplished by massed practice in a single session; the criterion for mastery was 10 correct choices in a row. Retention was measured by relearning (RL) of the critical habit under exactly the same procedures and to the same criterion. A trial was initiated by placing *S* in the start chamber, where he was detained until oriented towards the stem of the maze. A choice was not registered until his body had passed through the goal-box door, except for the tail. When this occurred the door was immediately closed, preventing him from retracing. The *S* was left in the goal box for at least 10 sec., or until he had consumed the three reward pellets on correct choices. He was then replaced immediately in the start box for another trial.

The position reversal occurred for experimental groups (two habits) without interruption once they had achieved criterion on the first task. Control animals (one habit) received somewhat different treatment depending upon the paradigm to which they had been assigned. PI controls were removed from their home cages approximately 30 min. before acquisition training; during this detention period they were housed in a small cage and given access to reward pellets commensurate with those earned by experimental animals on their first-habit acquisition. RI controls underwent analogous procedures after acquisition of their single habit. Immediate retention groups were tested without interruption after the second acquisition in the case of experimental animals, or, at corresponding time intervals in the case of controls. The other retention intervals were based on the time of reaching criterion on the to-be-tested habit (±10 min. for the 1-hr. group, ±1 hr. for the 1 day group, and ±3 hr. for the 1 wk. group).

RESULTS

Acquisition.—Comparability among the 16 groups was evaluated by a three-way factorial analysis of variance of trials to criterion on the initial habit (for experimental *S*s) or on the single habit learned by controls. All *F* ratios were equal to or less than 1.00 and thus there was no basis for denying equal learning ability among groups. For these data the overall mean number of trials to criterion was 16.75 with a range of 1–41 trials. For experimental *S*s only, the first task required a mean of 16.20 trials and the second required 18.00; however, positive and negative transfer were about equally likely (33 cases of positive transfer, 30 cases of negative transfer, and one tie). The correlation between first- and second-task acquisition speeds among experimental *S*s was +.11, which falls considerably short of statistical significance for a sample size of 64.

Retention.—The retention data are displayed in Fig. 1 according to four

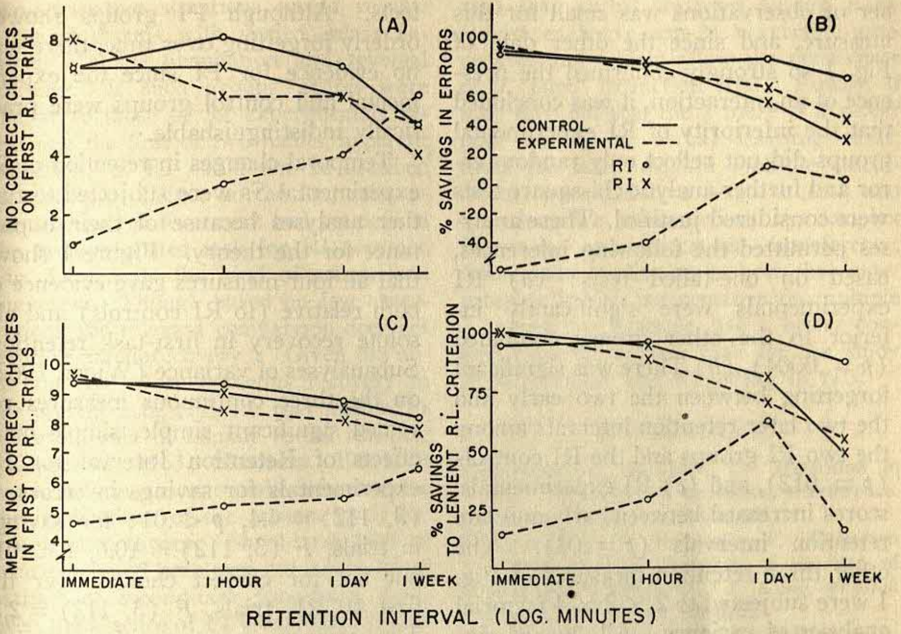


FIG. 1. Four measures of RL performance as a function of retention interval. (The parameter is treatment combination of Interference, control vs. experimental, and Paradigm, RI vs. PI.)

measures of performance. Figure 1a shows the number of correct choices made on the first relearning (RL) trial, or, since each *S* yielded only one observation for this measure, the number of *S*s (out of eight) whose first RL trial was a correct choice. The measure in Fig. 1c, is the mean number of correct choices on the first 10 RL trials. Figures 1b and 1d are based on savings scores, expressing RL performance in terms of an *S*'s OL performance. The former is the traditional Ebbinghausian measure calculated in terms of errors (i.e., OL errors minus RL errors as a proportion of OL errors). The last measure, in Fig. 1d, is a similar savings score, based on trials to criterion where a lenient RL criterion (9/10 correct) was used for computational purposes.

The first-trial data from RL (Fig. 1a) could not be subjected to the usual analysis of variance because they were based on a dichotomous measure and consequently violated the assumption of normality. A series of chi-square tests was therefore performed in the attempt to verify statistically what appeared obvious in Fig. 1a, viz., that there was an interaction between Paradigm and Interference caused by the marked inferiority of RI experimentals to the other groups and that the RI experimentals enjoyed a reduction in RI between immediate and 1-wk. tests. Winer's (1962) procedures for partitioning degrees of freedom and computing expectancies for multiple contingency tables under the assumption of no interactions yielded all nonsignificant terms, with the exception of the Paradigm × Interference interaction which was marginally significant, χ^2 $(1) = 2.59$, $p = .06$. Since the number of observations was small for this measure, and since the other data of Fig. 1 so strongly confirmed the presence of an interaction, it was concluded that the inferiority of RI experimental groups did not reflect only random error and further analytic chi-square tests were considered justified. These analyses permitted the following inferences, based on one-tailed tests: (*a*) RI experimentals were significantly inferior to the other groups combined ($p = .0004$), (*b*) There was significant forgetting between the two early and the two later retention intervals among the two PI groups and the RI controls ($p = .012$), and (*c*) RI experimentals' scores increased between early and late retention intervals ($p = .04$). The other three retention measures of Fig. 1 were subjected to $2 \times 2 \times 4$ factorial analyses of variance. All showed significant main effects of Paradigm; F $(1, 112) = 37.4$, 20.9, and 9.4, respectively, for the measures of Fig. 1c, 1b, and 1d, p's all $< .01$. Equally significant in all cases were the main effects of Interference; F $(1, 112) = 61.7$, 36.7, and 32.3, p's all $< .01$; and the interactions of Interference × Paradigm; F $(1, 112) = 47.7$, 45.5, and 26.3, p's all $< .01$. The main effect of Retention Interval was significant for savings in trials (Fig. 1d); F $(3, 112) = 4.8$, $p < .01$, but not for either of the other measures. The only other statistically significant effects were the Paradigm × Retention Interval interactions; F $(3, 112) = 3.7$, 5.4, and 4.0, p's $< .05$, .01, and .01, respectively. Figure 1 suggests that the marked inferiority of RI experimentals was responsible for the large Paradigm × Interference interactions and probably responsible also for the main effects of these two factors, since comparisons among the other three groups in some cases reverse the direction of such effects. Although PI groups showed orderly forgetting over time, there was no evidence for PI since the experimental and control groups were practically indistinguishable.

Temporal changes in retention of RI experimental *S*s were subjected to further analyses because of their importance for the theory. Figure 1 shows that all four measures gave evidence of both relative (to RI controls) and absolute recovery in first-task retention. Subanalyses of variance (Winer, 1962) on the three continuous measures revealed significant simple, simple main effects of Retention Interval for RI experimentals for savings in errors, F $(3, 112) = 4.4$, $p < .01$; for savings in trials, F $(3, 112) = 10.6$, $p < .01$; but not for correct choices over the first 10 RL trials, F $(3, 112) = 2.3$. The savings measures thus showed statistically significant recovery, complete after 1 day, while the other measures showed monotonic recovery up to 1 wk., but significant only for the first-trial choices.

DISCUSSION

The present experiment confirmed earlier data showing substantial RI among infrahumans but no PI (Kehoe, 1963; Rickard, 1965). Furthermore, in agreement with the predictions of interference theory (Postman, 1961), there was significantly better performance by RI experimentals at late than at early retention intervals. This time-related reduction in RI may not, however, automatically be identified as spontaneous recovery of first-task responses, since with only two possible responses recovery of one alternative yields the same data as would forgetting of the other, antagonistic response. Given the assumption that the first task is completely destroyed during acquisition of the second task, forgetting of the latter would look identical to recovery of the former. A definitive answer to the problem of such pseudorecovery in two-alternative situations would require a set of groups with varied delays between a single acquisition and reversal training. If such groups' scores were similar to those of RI experimentals relearning the first of two habits, apparent reductions in RI could not be ascribed to recovery. (Since on the first relearning trial no reinforcement had yet defined the to-be-relearned task for RI controls, Fig. 1a permits the kind of comparison suggested. Though based on few observations, the relevant comparison does not indicate pseudorecovery.) Given the uncertain role of spontaneous recovery in the present data, and, for that matter, its uncertain role in human verbal learning (Koppenaal, 1963; Slamecka, 1966), it would be premature to make inferences on the implications of the present study for interference theory. Should the kind of recovery found here turn out not to be an artifact of second-task forgetting, however, the absence of PI would seem highly paradoxical to the theory.

Purely at the empirical level, the present failure (and that of other investigators) to find effects of prior learning on retention presents a serious incompatibility between memory experiments on humans and infrahumans. Considerations of parsimony demand reluctance to accept species differences as being responsible for this incompatibility; therefore, several classes of task and procedure variables should be examined for their possible relation to PI. Distributed practice was one procedure identified by Kehoe (1963) as potentially responsible for her failure to find PI. However, the present experiment showed that massed practice in acquisition is not a sufficient condition for proactive effects. On the other hand, the present study used noncorrection procedures, as did Kehoe's. Since Gleitman and Jung (1963) obtained PI when using correction procedures, future research may reveal a critical importance of this variable, especially when it is appreciated that verbal learning studies make almost exclusive use of correction procedures and seldom allow the occurrence of a "pure" extinction trial. Finally, it may turn out that the T maze is a rather poor analog to verbal learning tasks, since the latter contain multiple associations, nonexhaustive response sets, and are therefore characterized by complex intratask and extratask interference. Subsequent studies may show that more complex habits, performance on multiple T mazes, for example, are a better context than the present task for the study of interference and forgetting in rats.

REFERENCES

GLEITMAN, H., & JUNG, L. Retention in rats: The effect of proactive interference. *Science,* 1963, **142,** 1683–1684.

GLEITMAN, H., & STEINMAN, F. Retention of runway performance as a function of proactive interference. *J. comp. physiol. Psychol.,* 1963, **56,** 834–838.

GLEITMAN, H., STEINMAN, F., & BERNHEIM, J. W. Effects of prior interference upon the retention of fixed-interval performance in rats. *J. comp. physiol. Psychol.,* 1965, **59,** 461–462.

KEHOE, J. Effects of prior and interpolated learning on retention in pigeons. *J. exp. Psychol.,* 1963, **65,** 537–545.

KOPPENAAL, R. J. Time changes in the strengths of A-B, A-C lists; spontaneous recovery? *J. verbal Learn. verbal Behav.,* 1963, **2,** 310–319.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.) *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152–179.

RICKARD, S. Proactive inhibition involving maze habits. *Psychon. Sci.,* 1965, **2,** 401–402.

SLAMECKA, N. J. Supplementary report: a search for spontaneous recovery of verbal associations. *J. verbal Learn. verbal Behav.,* 1966, **5,** 205–207.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received June 10, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 172–177

UNLEARNING IN THE A-B, A-C PARADIGM AS A FUNCTION OF PERCENTAGE OCCURRENCE OF RESPONSE MEMBERS[1]

GEOFFREY KEPPEL, BONNIE ZAVORTINK
University of California, Berkeley

AND BERNARD B. SCHIFF
University of Toronto

Manipulation of the percentage of occurrence of response members (ORM) in paired-associate (PA) learning has been equated with the operations defining partial reinforcement. The present studies tested the hypothesis that PAs acquired under reduced ORM would be more resistant to unlearning than PAs acquired under continuous ORM. The A-B, A-C paradigm was employed, with ORM being varied (50% vs. 100%) independently in both lists. The main difference between Experiments I and II was the reduced degree of A-B learning in Experiment II. While ORM produced only small differences in A-B learning, the variable greatly influenced A-C performance, an effect which was attributed to reduced within-condition variability and increased differences in the ORM effect. 1st-list recall was unaffected in Experiment I and only slightly related to 1st-list ORM in Experiment II. The importance of these findings for the ORM-partial reinforcement analogy was discussed, and it was concluded that the phenomena associated with these manipulations as well as their interpretations differ in the two situations.

Some theoretical interpretations in verbal learning have taken as a departure the assumption of a close correspondence between classical conditioning and paired-associate (PA) learning, equating the stimulus member of a pair with the CS, the presentation of the correct response term with the UCS, and the anticipated response member with the CR. Following the spirit of this equation, various investigators have attempted to extend the analogy by manipulating variables in PA learning which are known to influence classical conditioning. An example of this extension is the attempt to demonstrate the occurrence of spontaneous recovery following the unlearning of first-list responses during second-list learning (e.g., Koppenaal, 1963). More directly related to the problem under study in this paper, however, are the studies investigating the role of partial reinforcement in PA learning. In these experiments, partial reinforcement has been manipulated through a variation in the percentage of occurrence of response members (ORM). One implication of the equation with partial reinforcement is the expectation of an inverse relationship between ORM and retention. Specifically, under the assumptions that unlearning (*a*) represents an extinction-like process and (*b*) constitutes an important source of forgetting, it has been predicted that PAs acquired under reduced ORM would be forgotten at a slower rate than PAs acquired under continuous ORM. In one study, for example,

---

[1] This research was supported by a grant (MH-10249) from the National Institutes of Health. The paper was written at the Institute of Human Learning which is supported by grants from the National Science Foundation and the National Institutes of Health.

Schulz and Runquist (1960) compared the recall of lists 24 hr. following their learning under several percentages of ORM (20%, 40%, 60%, 80%, and 100%). In a similar experiment, Goss and Nelson (1962) compared the recall of lists learned under two schedules (25% and 100%) after 15 min., 24 hr., and 48 hr. Contrary to expectations, however, neither experiment showed any relationship between forgetting and ORM during acquisition.

It is clear that the failure of ORM to affect retention may have been due to the inappropriateness of one or more assumptions, e.g., the extinction-unlearning equation, or the ORM partial reinforcement analogy, or the assumption that forgetting is largely produced by unlearning. While not allowing a decision between these alternatives, the present studies extended the manipulation of ORM to a situation in which unlearning is known to occur, viz., the A-B, A-C paradigm. Although this experiment was originally suggested by Schulz and Runquist (1960), a search of the literature does not reveal any published studies concerned with this problem. In the two experiments to be reported, both first- and second-list ORM (50% and 100%) were varied independently. Experiment I was conducted with a relatively high degree of A-B learning in an attempt to equate potential differences in degree of learning under the two ORM conditions. However, since ORM did not affect List-1 recall, it was considered possible that the high degree of learning might have obscured relatively small differences in the amounts of unlearning. Thus, in Exp. II, the degree of first-list learning was reduced drastically and two ORM control groups were included to provide an estimate of possible differences in degree of first-list learning.

METHOD

Design and procedure.—The basic design of both experiments consisted of a factorial combination of first- and second-list ORM (50% and 100%). For Exp. I the first list was continued for an additional five trials following the attainment of a perfect recitation; the second list was administered for 15 trials. The anticipation method was used with the lists being presented on a memory drum at a 2:2-sec. rate. The intertrial interval was 4 sec. Immediately following training, *S* was asked to recall the first list. The recall test corresponded to the one used by Barnes and Underwood (1959), except that second-list responses were provided, unpaired, on the recall sheet. A 2-min. time limit was imposed. While *S*s were asked to pair correctly recalled responses, they were also encouraged to guess or to list responses unpaired at the bottom of the recall sheet. A similar general procedure was followed in Exp. II, except that the first list was taken to a criterion of 8/12 correct responses and the recall test was administered on the memory drum. On this test, recall was self-paced and *S* was provided with correctly paired second-list responses. Control *S*s were given an arithmetical reasoning test which was continued for a time equivalent to that required for the 15 interpolated learning trials.

Materials.—In both experiments, the two lists corresponded to the A-B, A-C paradigm. The lists of eight CVC adjective pairs, employed by Barnes and Underwood (1959), were used in Exp. I. There were two different pairings of the syllables and adjectives and the two lists were presented in the same order to all *S*s. The pairs were arranged in four orders with the restrictions that no pair appear in the same serial position or in the same sequential relationship with another pair more than once. For 50% ORM, four CVCs were paired with responses and four were followed by a blank on each trial. The particular responses that were seen in Orders 1 and 3 were independently randomly determined, while Orders 2 and 4 contained the sets of four responses complementary to 1 and 3. The *S*s were started on Order 1 or 3, so that each *S* saw all eight responses in the first two presentation trials and for all successive blocks of two trials.

The lists in Exp. II consisted of 12 two-syllable adjective pairs, previously employed by Postman (1966). Each list was arranged into two independent pairings and occurred first equally often. As in Exp. I, four different serial orderings of the pairs were used. For Exp. II, however, the ORM schedule was balanced in blocks of four, with the restriction that six responses and six blanks appear on each trial. Thus, for the 50% condition, each response was presented twice within blocks of four trials. At recall, the stimuli were presented in a new random ordering.

Subjects.—The *S*s in both experiments were drawn from psychology classes at the University of California. There were 16 *S*s in all experimental groups and 8 *S*s in each of the two control groups used in Exp. II. The four main conditions of Exp. I were randomized in blocks of four with list pairings and starting orders being balanced randomly over 16 blocks. A similar arrangement was constructed for Exp. II, except that control conditions were randomized within successive blocks of two. The *S*s were assigned to conditions in order of appearance. One *S* was dropped from Exp. II for failure to reach criterion.

RESULTS

List-1 learning.—Criterial scores failed to show reliable differences between the two ORM conditions. For Exp. I, the average numbers of trials to reach a criterion of 8/8 correct responses were 9.16 and 8.81 for Cond. 100 and 50, respectively, $F < 1$. For Exp. II, 6.88 and 8.38 trials were required to reach a partial criterion of 8/12 correct responses, F (1, 74) = 2.91, $p > .05$ (following a logarithmic transformation). None of the remaining sources of variance was significant in the two experiments.

TABLE 1

MEAN NUMBER OF CORRECT RESPONSES ON TRIAL 1

	Exp. I		Exp. II	
	100%	50%	100%	50%
List 1				
Seen	.94	2.28	1.25	2.41
Not Seen	.84	.00	1.53	.00
List 2				
Seen	.81	1.00	.94	2.09
Not Seen	.97	.00	.72	.00

While ORM did not influence overall rate of learning, it is still of some interest to examine the influence of this variable on the initial learning trials. One such analysis consisted of separation of pairs for which the correct response was either "seen" or "not seen" on the initial study trial. The results of this tabulation for Trial-1 performance are presented in the upper half of Table 1. The values listed for Cond. 100 are based upon the same pairs associated with Cond. 50. For both experiments, the results were similar, the mean number of recalled pairs, which were first seen on the initial study trial for Cond. 50, is approximately twice as large as that obtained for equivalent pairs presented under 100% ORM. This superiority was maintained over the first four trials, after which time ceiling effects (Exp. I) and *S* dropouts (Exp. II) precluded an extension of the analysis beyond Trial 4.

List-2 learning.—Performance over the 15 interpolated trials was largely determined by the List-2 ORM schedule. The average numbers of correct anticipations for second-list 100% and 50% conditions, respectively, were 95.50 and 79.81 in Exp. I, F (1, 60) = 16.21, $p < .01$, and 128.91 and 113.62 in Exp. II, F (1, 60) = 4.60, $p < .05$. Neither first-list ORM nor the change in ORM schedule affected List-2 performance (all Fs < 1). Similar results were found with the criterial measures, 6.96 and 9.84 trials required to reach perfect performance in Exp. I for Cond. 100 and 50, respectively, and in Exp. II, 4.97 and 7.47 trials for a partial criterion (8/12 correct responses), F (1, 60) = 13.51 and 8.48, $p < .01$.

An analysis of the effect on Trial-1 performance of initial exposure of the correct pairing on the study trial can be found in the lower portion of Table 1. This time the data from the two experiments are not in agreement, Exp. I showing roughly equivalent performance for "seen" pairs and Exp. II revealing the nearly doubled performance previously shown for the first list in both studies. On subsequent trials, however, the two experiments are in substantial agreement with the first-seen pairs showing marked inferiority under 50% ORM. This finding is in contrast to the relatively consistent superiority for Cond. 50 of the first-seen pairs over the early List-1 trials.

Recall.—The mean numbers of first-list responses which were correctly paired are presented in Table 2. In Exp. I, slightly poorer recall was obtained for first-list Cond. 50, but neither this difference, second-list ORM, nor the interaction proved to be significant (all Fs < 1). Although there was no RI rest control, it is obvious that unlearning must have been produced since the average recall of the experimental conditions, 6.11, is considerably lower than the perfect score (8.0) which most likely would have been approximated under control conditions.

The recall scores for Exp. II were subjected to a similar analysis. A comparison of experimental and control conditions indicated that unlearning did occur, $F\ (1, 74) = 30.51$. With this design, the appropriate test of the effect of List-1 ORM on the magnitude of the unlearning effect is the interaction of List-1 training with the experimental/control dichotomy. This comparison, while in the direction predicted by the ORM partial reinforcement analogy, failed to reach significance, $F < 1$. Other methods of scoring, e.g., total recall regardless of pairing, produced identical results. On the other hand, it is true that the recall results are qualitatively in the expected direction for both sets of List-1 ORM conditions. Considering these two comparisons (100–50 vs. 50–50; 100–100 vs. 50–100) as independent tests of the same hypothesis shows this consistency to be reliable, $\chi^2(4) = 10.80$, $p < .05$ (Winer, 1962, pp. 43–45). In view of these marginal results, the most reasonable conclusion would be that first-list ORM produced a small, but consistent effect and that perhaps with a still lower degree of first-list learning and a further reduced ORM (e.g., 25%), the magnitude of the relationship would be greater.

TABLE 2

MEAN NUMBER OF LIST-1 RESPONSES CORRECTLY RECALLED AND PAIRED

List-2 ORM	Exp. I		Exp. II	
	List-1 ORM		List-1 ORM	
	100%	50%	100%	50%
100%	6.12	6.12	5.88	7.43
50%	6.44	5.75	6.06	6.43
Combined	6.28	5.94	5.97	6.93
Control	—	—	9.38	9.50

While the interaction of the two training variables is not significant ($p > .10$), the patterning of the means in both experiments is such that greater unlearning occurs for *S*s receiving the same ORM schedule on the two lists than for *S*s receiving different schedules. These findings suggest that method change can produce a decrease in the unlearning effect and are consistent with the findings of Postman, Keppel, and Stark (1965) who found a larger, but similar effect for change of response-term form class.

DISCUSSION

In neither of the two experiments did ORM significantly influence first-list performance. However, since the response terms in both studies consisted of adjectives, this finding does support the generalization that the magnitude of the ORM effect is inversely related to the meaningfulness of the responses (e.g., Schulz, 1965). In marked contrast to the negative first-list results, however, ORM greatly influenced second-list performance. This state of affairs appears to have been brought about by two contributing factors, (*a*) reduced within-condition variability, and (*b*) increased differences in the ORM effect. With respect to the first factor, the within-mean squares were at least halved in both studies. Support for the second factor is the numerical increase in the ORM effect on the second list. While neither interaction is reliable, the consistency of the effect does suggest that the influence of ORM tends to increase with training. It is not clear, though, whether this is to be attributed to an increase in associative interference for Cond. 50 or to a nonspecific transfer effect. This question could be answered through the comparison of transfer-list performance with the A-B, A-C and A-B, C-D paradigms.

The acquisition data of the present experiments and those reported by others suggest that while ORM can be shown to be an effective variable in PA learning, ORM phenomena tend to be qualitatively different from partial-reinforcement phenomena. For example, several investigators have commented upon the fact that ORM effects are much smaller than those found in conditioning studies (cf. Goss, Morgan, & Golin, 1959; Schulz & Runquist, 1960). In addition, there are variables which interact with ORM, but which may be difficult to translate into the conditioning paradigm, e.g., response-term meaningfulness. More importantly, though, is the fact that the interpretation of ORM phenomena tends to be largely in terms of constructs and processes which have been developed primarily in the area of verbal learning, rather than in the area of classical conditioning. For instance, in order to account for the smaller ORM effect noted above, Goss et al. suggested several processes operating in the PA situation which might tend to counteract the potentially detrimental effects of ORM, e.g., reduced response competition, increased opportunity for rehearsal, introduction of time interval between particular PAs, and increased stimulus familiarization. Certainly the first two factors, for which empirical support has been given by Schulz and Runquist (1960), only make sense in the PA paradigm. Finally, the analysis by Schulz (1965) of the ORM effect into response- and associative-learning stages represents an interpretation which may be difficult to apply to conditioning studies. In short, the analogy between ORM and partial reinforcement is more apparent than real, with the phenomena as well as their interpretations differing in the two situations.

The main purpose of this study, however, was to determine whether PAs learned under partial ORM would be more resistant to unlearning than PAs learned under continuous ORM. The unlearning of List-1 responses was unaffected in Exp. I, and only slightly related to first-list ORM in Exp. II. This small discrepancy between the two experiments may be due to the reduced degree of first-list learning employed in Exp. II. On the other hand, it may be that this effect of ORM is the result of an artifact produced by differential first-list overlearning in Cond. 50. That is, to the extent that more trials are required to reach List-1 criterion under 50% ORM, some of the easier pairs will receive greater overlearning than corresponding pairs learned under 100% ORM. Although an analysis of easy-difficult pairs in the present experiments did not reveal an interaction of difficulty and first-list ORM, the magnitude of this potential artifact should increase with the effect of ORM on acquisition. Thus, for example, differential overlearning should become a problem with materials of low meaningfulness used as response terms or with lower percentages of ORM—any procedure which would tend to magnify the List-1 ORM effect.

On the basis of the results of this and other studies, it must be concluded that ORM does not influence forgetting and only slightly influences unlearning in the A-C paradigm. While these findings suggest that extinction and unlearning are not functionally equivalent, it could be argued that the failure to find similar phenomena in the two situations is due to the fact that truly analogous experiments have not been conducted. Moreover, it is not immediately apparent whether such a comparison study could be conducted. For instance, in order to insure the presence of stimulus specificity, as is the case in a minimal PA experiment, it will be necessary to employ two CS-UCS pairs for original and interpolated learning. In addition, the UCRs must be unrelated in the sense that the initial associative strength of second-list pairs must have been unaffected by first-list acquisition. Otherwise, the transfer relationship would correspond more to an A-B, A-B′ paradigm (where B and B′ are associatively or semantically related) than to an A-C paradigm. Provided this analysis is correct, it would appear that a detailed comparison of extinction as studied in classical conditioning and unlearning as studied in the A-C paradigm has yet to be performed.

REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.,* 1959, **58,** 97–105.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *J. exp. Psychol.,* 1959, **57,** 96–104.

GOSS, A. E., & NELSON, J. V. D. Retention of paired associates as a function of percentage of occurrence of response members. *J. gen. Psychol.,* 1962, **66,** 203–211.

KOPPENAAL, R. J. Time changes in the strengths of A-B, A-C lists; spontaneous recovery? *J. verbal Learn. verbal Behav.,* 1963, **2,** 310–319.

POSTMAN, L. Differences between unmixed and mixed transfer designs as a function of paradigm. *J. verbal Learn. verbal Behav.,* 1966, **5,** 240–248.

POSTMAN, L., KEPPEL, G., & STARK, K. Unlearning as a function of the relationship between successive response classes. *J. exp. Psychol.,* 1965, **69,** 111–118.

SCHULZ, R. W. Learning of paired associates as a function of pronunciability and percentage occurrence of stimulus or response members. *J. verbal Learn. verbal Behav.,* 1965, **4,** 494–497.

SCHULZ, R. W., & RUNQUIST, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *J. exp. Psychol.,* 1960, **59,** 409–413.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received June 13, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 178–181

STIMULUS VARIATION AND REPETITION IN THE ACQUISITION OF NAMING RESPONSES[1]

WILLIAM F. DUKES AND WILLIAM BEVAN
University of California, Davis *Johns Hopkins University*

A total of 257 female Ss learned to associate names with photographs, each S serving under 1 of 2 training and 2 test conditions. In 1 training condition, the same photograph of each person was repeatedly presented; in the other, 4 different photographs were used. In 1 test situation, the test photos were a set that had been used in training; in the other, they were completely new poses. 3 different levels of training were used. Results indicate that repetition of particular S-R combinations is the important training principle when the purpose of training is the reinstatement of a particular R to a specific individual S; in contrast, stimulus variation is the significant principle when the purpose is correctly identifying members of a generic class. The significance of these 2 principles was clarified by comparing performance across different levels of training. A general practice effect was also found to be associated with the use of stimulus variation.

In a previous paper the present authors (Bevan, Dukes, & Avant, 1966) demonstrated that variation among instances of a common categorical stimulus increases the probability of its recall. The data indicated that x will be better remembered if it occurs once as x_1 and later as x_2 than if either x_1 or x_2 occurs twice. The present paper raises the question of how variation operates in conjunction with frequency in the process of learning a particular response to one or more of a family of stimuli. Specifically it is asked, "Will R be more readily evoked when training has been of the $S_A - S_B - S_C - S_D$ or the $S_A - S_A - S_A - S_A$ type?" This general question is, of course, not new, having often been under scrutiny in research on problem-solving and concept-formation behavior—recently by such investigators as Lloyd (1960), Battig and Bourne (1961), Smith, Jones, and Thomas (1963), and Petre (1966). Although the importance of this type of stimulus variation in concept learning has been recognized (cf. James, 1890), findings about its effectiveness have not been wholly consistent. Data provided by Adams (1954), e.g., indicate that single-problem training is superior to multiple-problem training in human problem solving, while Callentine and Warren (1955) found the reverse, explaining the discrepancy on the basis of operational differences.

Any attempt to answer the question as posed in the preceding paragraph entails first answering the implicit question, "Evoked by what?" The relative efficiency of the two types of training will undoubtedly differ depending upon whether the "evoking" stimulus (that used in the posttraining procedure) is a new member (S_E) or a particular one previously presented during training (S_A). Consequently, both types of test stimuli are applied in the present work.

---

[1] This research was conducted while W. B. was in residence as a National Science Foundation Fellow and Center Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.

METHOD

Subjects.—Members of undergraduate psychology and education classes at the University of California, Davis, served as *S*s. Since the scores of male *S*s, although consistently lower, manifest the same trend as those of female *S*s and since the proportion of males to females varied widely among the various experimental groups, the analysis and discussion include only the records of the latter ($N=257$).

Materials.—The stimulus materials were color photographs of 20 persons, 10 male and 10 female, cut from mail-order catalogs; the photographs were primarily head-to-waist extensions but included some head-to-knee shots. From a selection of different garbs—e.g., beachwear, formal dress, etc.—and poses—full face to 45° oblique—of each person, five sets of 20 photographs (labeled "A" through "E") were assembled so as to maximize intraset variation. Every set contained one photograph of each person, but no exact duplications occurred across sets. For example, even though the person given the name "Frank Oliver" appeared in all five sets, he was dressed in a blue suit in Set A, a bathrobe in Set B, etc. The photographs in each set were individually mounted on sheets of colored paper; the color was constant for all members of a given set, but varied from set to set. Short, common American names were assigned each of the 20 persons and, except in the test series, were typewritten on strips of white paper mounted above the head. The same name was, of course, used in the several training presentations of a particular person. For example, whether in a blue raincoat, a pink sweater, or otherwise clothed, the girl assigned the name "Kay Wagner" was so identified throughout the training series. Slides were made of each picture for projection in a Kodak Carousel 35-mm. projector.

Procedure.—The data were collected in groups which ranged in size from 2 to 50 *S*s. Directions were simple and essentially the same for all groups. Each group was told that they would see on the screen before them pictures of 20 different persons identified by name, and that although individual presentations would be relatively brief, each person would appear several times. (In one control group each person appeared only once, and appropriate modifications were made in the instructions.) The *S*s were requested not to talk or make notes while the slides were being projected but merely to attend the screen carefully, observing the person-name pairings. Next they were informed that following the training series they would be shown the persons without the names and asked to identify them.

During training (name-face presentations) each slide was shown for 5 sec. Upon completion of training, *S*s were given sheets of paper with spaces numbered 1 through 20 and told that they now would see slides of the same 20 people again but without the names. No statement was made about whether the test stimuli were new or repeated versions of those shown during training. In all groups *S*s were simply requested to write the name of each person in the proper blank as the unidentified photographs were projected. The 20 pictures without names were then shown for 15 sec. each.

Design.—A balanced design with two conditions of training and two types of test situation was used. In one training condition (F) the same set of photographs—Set A, Set A, etc.—was repeatedly presented. In the other training condition (V) four different sets of photographs—sets A, B, C, D—were used. In one test situation (f) the photographs were a set—Set A—that had been used in the training series; in the other (v), completely new poses—Set E—were presented. Different within-set random orders were used for Sets A through D, and these orders were duplicated in the four presentations of Set A. Degree of training was varied simply by presenting the repeated set 4, 8, or 16 times and the four-set complex, 1, 2, or 4 times. Thus, in all 12 independent experimental groups were used; but because testing was conducted in the classroom or in the laboratory with volunteers, the number of *S*s serving under each condition varied from 8 to 48.

In addition, for comparison purposes two control groups not used in the overall analysis were tested. One viewed the set used in repeated training (A) only one time, the other twice, and both were tested on the photographs used in training.

RESULTS AND DISCUSSION

With the double-name identification a number of alternative definitions of a correct response are obviously available. A strict criterion would allow credit only when both first and last names are given; a lenient one would accept one name, either first or last (or both), as adequate. A compromise

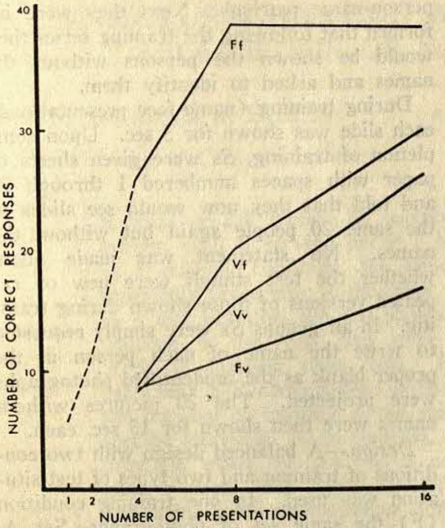


FIG. 1. Recall scores for *S*s trained to associate names with photographs under each of the experimental conditions. (F indicates that the same photograph was always paired with a particular name during training; V indicates that the person named appeared in different garb and poses on repeated trials; f indicates that the test photograph was one used in training—A; v indicates that a new pose and garb were used in the test photo —E.)

might score two points for each correct two name and one point for each correct single name identification of the test stimuli whether they were those presented in training—Set A—or new versions of these—Set E. Under such a system a maximum score of 40 is possible.

Although each of these three measures was applied to *S*s' responses, discussion is restricted to the data derived from the last mentioned since the same basic trends are exhibited regardless of which measure is used. The compromise was selected primarily because it recognizes partial identification while weighting full naming more heavily.

Mean scores for the various groups are presented graphically in Fig. 1. An analysis of variance was performed on the data, appropriate corrections being made for unequal cell frequencies.

The efficacy of the two types of training clearly differed with the type of stimulus used to test recall. When the test stimulus had been presented as a training stimulus, repetition in training yielded higher recall scores than variation—cf. the curves marked Ff and Vf, F (1, 222) = 135.74, $p < .001$, for the mean difference in performance under these two conditions. When, on the other hand, the test stimuli were different from, but related to those used in training, the advantage of stimulus variation is apparent—cf. the Fv and Vv curves. Although for four presentations no difference in mean performance under these two conditions existed, for both 8 and 16 presentations mean differences are significant, t_8 (41) = 2.23, $p < .05$, t_{16} (22) = 3.85, $p <$ .01.

Thus the superiority of intensive drill on particular S-R combination is clear when the purpose of training is to reinstate a particular R to a specific individual S, while the importance of diverse contacts with members of a class of stimuli is obvious when the purpose of training is to correctly identify new members of the class.

Since the test stimuli in the v condition are in one sense previously presented stimuli—they are the same people—and in another, novel presentations, alternative interpretations may be placed on the data obtained from it. If one emphasizes the former (the sameness aspect), the results may be discussed in terms of recall. If, on the other hand, the newness of the test stimuli is stressed, correct responses are appropriately treated as generalizations. According to this latter interpretation the superiority of Vv over Fv could be attributed to a "set to generalize" which is more effectively generated by varied training.

Basic to generalizations about the relative merits of the two training conditions is a consideration of the general level of development of the S-R system. At a low level of acquisition (i.e., performance after four presentations) no differences are observed between any combination of Fv, Vf, and Vv; only Ff is clearly differentiated from the others. In contrast at higher levels of acquisition (after 8 or 16 presentations) differences in the performance levels of the several groups are readily apparent. The higher initial level of performance in the Ff group suggests the propaedeutic role of repetition in establishing a core of particular responses from which systems of categorical responses may evolve.

The inability to secure a larger number of different photographs of each person imposed an obvious restriction on the number of variations per stimulus. Had it been possible to match the 8 or 16 repeated photographs with an equal number of varied presentations (instead of four variations repeated), the differences among the experimental conditions may have been more striking.

As might be expected, the varied practice procedure yields what might be termed a "general practice effect." Four varied presentations involve only one presentation of a particular stimulus and eight presentations only two, yet the mean scores under these conditions are higher when tested for identification of the stimulus used in training (8.73 and 20.05, respectively) than under comparable conditions of repeated training (5.69 and 10.60). The t's for the mean differences are: one presentation, $t\ (37) = 2.49,\ p < .05$; two presentations, $t\ (28) = 3.17,\ p < .01$. The magnitude of this effect presumably depends upon the degree of similarity among the members composing a family of stimuli.

Finally, the relative difficulty of acquiring class responses as compared to that of acquiring simple S-R relationships is indicated by the performance levels in the Vv and Fv vs. the Ff and Vf experimental conditions, $F\ (1, 222) = 186.84,\ p < .001$ for the two test conditions, f and v. These findings support those from Metzger's (1958) comparison of rote learning and concept formation in which doubling the number of stimuli while keeping responses constant resulted in increased difficulty of learning and furthermore confirm Smith, Jones, and Thomas's (1963) conclusion that " . . . increases in the number of stimuli per response hinder rote learning but aid concept formation [p. 475]."

REFERENCES

ADAMS, J. A. Multiple versus single problem training in human problem solving. *J. exp. Psychol.*, 1954, **48**, 15–18.

BATTIG, W. F., & BOURNE, L. E., JR. Concept identification as a function of intra- and interdimensional variability. *J. exp. Psychol.*, 1961, **61**, 329–333.

BEVAN, W., DUKES, W. F., & AVANT, L. L. The effect of variation in specific stimuli on memory for their superordinates. *Amer. J. Psychol.*, 1966, **79**, 250–257.

CALLANTINE, M. F., & WARREN, J. M. Learning sets in human concept formation. *Psychol. Rep.*, 1955, **1**, 363–367.

JAMES, W. *Psychology*. New York: Holt, 1890.

LLOYD, K. E. Retention of responses to specific classes and to specific stimuli. *J. exp. Psychol.*, 1960, **59**, 54–59.

METZGER, R. A comparison between rote learning and concept formation. *J. exp. Psychol.*, 1958, **56**, 226–231.

PETRE, R. D. Effect of amount or pretraining with identical and dissimilar stimuli on concept learning. *J. exp. Psychol.*, 1966, **71**, 472–474.

SMITH, T. A., JONES, L. V., & THOMAS, S. Effects upon verbal learning of stimulus similarity, number of stimuli per response, and concept formation. *J. verbal Learn. verbal Behav.*, 1963, **1**, 470–476.

(Received June 15, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 182–186

TWO-CHOICE DISCRIMINATION LEARNING AS A FUNCTION OF CUE SIMILARITY AND PROBABILITY OF REINFORCEMENT[1]

JOSEPH HALPERN AND JOHN W. MOORE

University of Massachusetts

72 human Ss served in a two-choice auditory discrimination learning experiment in which 3 levels of cue similarity ($\Delta f = 0$, 20, 60 cps) were crossed with 2 noncontingent reinforcement schedules ($\pi_1 = 1$ and $\pi_2 = .5$; $\pi_1 = .9$ and $\pi_2 = .6$). Marginal response probabilities were reasonably consistent with generalization models (e.g., Burke-Estes, 1957) in being bounded by π values under low cue similarity and in converging toward the average π value as cue similarity increased. However, first-order conditional statistics from the last 100 trials were inconsistent with expectations from these and other reinforcement models.

The purpose of this experiment was to evaluate the joint effects of cue similarity and probability of reinforcement on two-choice discrimination learning. The situation resembled a typical probability learning task in which *S*s are instructed to predict which of two reinforcing lights (E_1 or E_2) will terminate a trial. The probability of E_1 given one of the two successively presented discriminative stimuli, $P(E_1|S_i) = \pi_i$ ($i = 1,2$), was not contingent on *S*'s choice response (A_1 or A_2). The discriminative stimuli (S_1 and S_2) were pure tones, and similarity was varied along the auditory frequency dimension.

Following Uhl (1964), an additional purpose of this study was to assess the effectiveness of the Burke-Estes (1957) linear model in describing the data. The Burke-Estes model is a stimulus sampling theory in which discriminative stimuli are treated as consisting of a large number of elements from which a random sample is drawn on each trial. Asymptotic response probability is a joint function of the proportion of elements common to the two stimuli and the π_i values. When the a priori probabilities of S_1 and S_2 are equal, the probabilities of an A_1 response given S_1 and S_2 are

$$P_{11} = P(A_1|S_1) = (1-w)\pi_1 + w\pi_a \quad [1]$$

$$P_{12} = P(A_1|S_1) = (1-w)\pi_2 + w\pi_a \quad [2]$$

where $\pi_a = (\pi_1+\pi_2)/2$, $w = N(S_1 \cap S_2)/N(S_1) = N(S_1 \cap S_2)/N(S_2)$, and where $N(\cdot)$ refers to the number of stimulus elements in a set. Equations 1 and 2 clearly predict probability matching, $P_{1i} = \pi_i$, whenever $w = 0$ and probability averaging, π_a, whenever $w = 1$. In fact, P_{1i} is a linear function of $0 \leq w \leq 1$ bounded by their two extremes.

Previous experiments have varied π_i or stimulus similarity, but not both. In experiments involving variation of π_i,

---

[1] This investigation was supported in part by Public Health Service Grant NIH-MH-10794-01. The authors wish to acknowledge the assistance of Douglas White.

[2] $$\chi^2(df) = \sum_{\substack{i,j,k,m \\ (df+1\text{ times})}}^{2} \frac{[P(A_1|S_iS_jA_kE_m) - P(A_1|S_i)]^2 N_{ijkm}}{P(A_1|S_i)}$$ where N_{ijkm} is the number of cases associated with each conditional.

Estes and Burke (1955) and Wollen (1963) provide data consistent with the linear model while Popper and Atkinson (1958) and Atkinson, Bogartz, and Turner (1959) do not. Uhl (1964) defined stimulus similarity as the proportion of overlap between two sets of lights. The results of the latter study presumably support a pattern model over the linear model in that there were no significant departures from probability matching with increases in cue overlap. Uhl points out that his experiment might have been biased against the Burke-Estes model because light matrices lend themselves to pattern rather than components discrimination. In fact, since there would be fewer lights to monitor in order to decide which stimulus is presented on a trial, increasing the proportion of lights common to S_1 and S_2 might actually emphasize the discriminative aspects of *S*'s task and thereby facilitate discrimination performance. Auditory cues were employed here in order to eliminate these complexities.

Method

Subjects.—The *S*s were 72 University of Massachusetts undergraduates distributed randomly into six groups of 12 *S*s each.

Apparatus.—Two *S*s were run at a time. Each *S* was seated before a small Estes-Straughan conditioning board consisting of a white warning light and two red event lights. Two spring loaded switches were mounted below each of the event lights. An opaque screen separated adjacent *S*s. Auditory stimuli produced by two Hewlett-Packard Model 200 audio oscillators were modulated by two attenuators and presented to *S*s over calibrated headphones. Experimental events were controlled by automatic programming equipment and responses were recorded on an Esterline-Angus event recorder. A continuous white noise of 60-db. SPL was used to mask irrelevant sounds.

Procedure.—The six groups of *S*s differed in the reinforcement schedule used, $\pi_1 = 1.0$ and $\pi_2 = .5$ or $\pi_1 = .9$ and $\pi_2 = .6$, and the frequency differential Δf which separated the two discriminative stimuli. The frequency pairings in cps were 740–800 ($\Delta f = 60$), 780–800 ($\Delta f = 20$), and 800–800 ($\Delta f = 0$). The intensity of all tones was 70-db. SPL.

Each *S* received a total of 200 trials, 100 trials to each stimulus. Each trial consisted of the presentation of one of the two tones for 5 sec., a warning signal light occurring 1.7 sec. after the onset of the tone, and a reinforcing light occurring 1.7 sec. after the onset of the warning signal, its duration being 1.6 sec. Total time per trial was 5 sec. and the intertrial interval was a constant 5 sec.

The sequence of the 200 trials was randomly chosen with the restriction that there be no more than four S_1 or S_2 trials in succession, and two different random sequences were employed for each reinforcement schedule. The effect of the position of the more frequently reinforced stimulus was counterbalanced by designating the right-hand response switch and light A_1 and E_1 for half of the *S*s in each group and the left hand and light A_1 and E_1 for the other half.

The *S*s were instructed to maximize correct responses. No mention was made of the discriminative aspects of the task.

Results and Discussion

Marginal statistics.—Observed values of P_{11} and P_{12} in blocks of 40 trials for each combination of Δf and π appear in Fig. 1. The graph shows that for $\Delta f = 60$, P_{11} and P_{12} fell within the bounds required by the Burke-Estes model. Furthermore, the functions converge nicely toward π_a with decreases in Δf. Analysis of variance based on the last 40 trials indicated that the effect of Δf on discrimination performance $(P_{11} - P_{12})$ was significant, $F\ (2, 48) = 11.41$, $p < .001$.

Figure 1 also shows that the rate of convergence of P_{11} and P_{12} toward π_a was faster with $\pi_1 = 1$ and $\pi_2 = .5$ than with $\pi_1 = .9$ and $\pi_2 = .6$; and, unexpectedly, discrimination performance under the latter reinforcement schedule was actually best at the intermediate level of cue similarity. This particular aspect of the interaction between cue similarity of reinforcement probability,

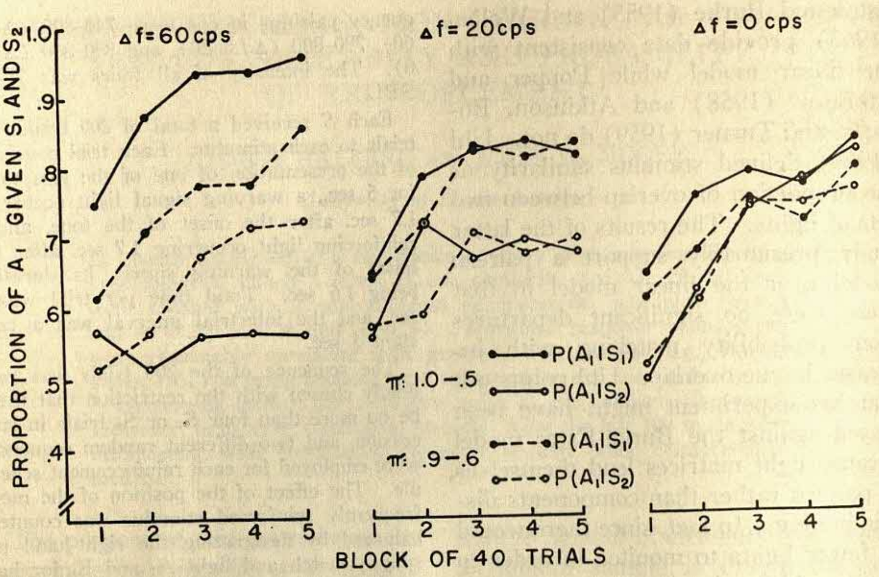


Fig. 1. Proportion of A_1 responses given S_1 and S_2 as a function of blocks of 40 trials for each combination of Δf and π.

$F(2, 48) = 7.21$, $p < .01$, is not predicted by the model.

Table 1 shows the interaction or lack of independence between cue similarity and reinforcement probability in a different way. Estimates of w in the table were obtained by solving Equations 1 and 2 using the observed values of P_{11} and P_{12} for each group. These estimates of w depended rather strongly on the values of π_i. This is especially clear for $\Delta f = 60$ where the similarity index $\hat{w}$ was .673 with $\pi_1 = .9$ and $\pi_2 = .6$ but only .264 with $\pi_1 = 1$ and $\pi_2 = .5$. One explanation may be that the perfect reliability of a stimulus associated with a π value of unity acts as an additional cue which can enhance discrimination performance and hence decrease estimates of the similarity parameter. However, this interpretation is contradicted by the comparatively small $\hat{w}$ of .520 for the $\Delta f = 20$ group where $\pi_1 = .9$ and $\pi_2 = .6$, but this particular inversion may well represent sampling error.

TABLE 1

Estimates of w Computed with Equations 1–2 from Observed Values of P_{1i} from Trials 101–200

Group	Stimulus Similarity (Δf)			
	60	20	0	Pooled
π:				
1.0–.5	.264	.724	.972	.653
.9–.6	.673	.520	.940	.711

Apart from the fact that $\hat{w}$ was a function of π, the Burke-Estes model describes the marginal statistics reasonably well. In fact, these data provide some support for the entire class of simple generalization theories, a class which also includes the Bush and Mosteller (1951) model and an interesting recent entry by Lee (1966). By contrast, the Popper and Atkinson (1958) experiment which employed visual cues and the Uhl (1964) experiment demonstrated inadequacies in the generalization hypothesis. Because of this, the earlier models were set aside in favor of more complex alternatives such as Atkinson's (1958) "observing response" model.

Sequential statistics.—Reinforcement theories, such as the Burke-Estes model and others previously mentioned, require that first-order conditional statistics of the form

$$P(A_{1,n}|S_{i,n}S_{j,n-1}A_{k,n-1}E_{m,n-1})$$

possess certain ordinal properties which can serve as a rough test. All of these models require that for each combination of successive trials, $S_{i,n}S_{j,n-1}$, the conditional $P(A_1|S_iS_jA_1E_1)$ be the largest of the set of four, $P(A_1|S_iS_jA_2E_2)$ the smallest, and that $P(A_1|S_iS_jA_1E_2)$ and $P(A_1|S_iS_jA_2E_1)$ fall somewhere between these two, depending on parameters of the model. This pattern has been verified reasonably often in simple nondiscriminative choice situations (Myers & Atkinson, 1965).

Table 2 shows that inversions of this pattern were common under both reinforcement schedules, and subsequent experiments in our laboratory have yielded similar results (Halpern, 1966; Moore & Halpern, 1967). Extensive inversions of the first-order conditional statistics have also been observed by Myers and his associates in within-*S*s comparisons of incentive effects (e.g., Lipkin, 1966). In these experiments, *S*s experience two types of trials or cues, each having its own independent π value and amount of reward for correct event prediction. Evidently, inversions in the expected ordering of conditional statistics can occur in a variety of successive-cue choice situations.

One might be tempted to interpret the data in Table 2 in terms of the possibility that, at least asymptotically, $P(A_1|S_i)$ on Trial n is *independent* of the events of the preceding trial. If this were the case, the conditional statistics would not differ significantly

TABLE 2

OBSERVED FIRST-ORDER CONDITIONAL PROBABILITIES OF A_1 RESPONSE ON TRIAL n FOR TRIALS 101–200 POOLED OVER LEVELS OF Δf

Trial n-1			Trial n Cue	Reinforcement Schedule	
Cue	Response	Event		$\pi_1=1$ & $\pi_2=.5$	$\pi_1=.9$ & $\pi_2=.6$
1	1	1	1	.872 (493)[a]	.831 (450)
1	1	2	1	—	.772 (123)
1	2	1	1	.710 (136)	.693 (75)
1	2	2	1	—	.727 (44)
2	1	1	2	.716 (278)	.660 (350)
2	1	2	2	.755 (141)	.804 (92)
2	2	1	2	.450 (118)	.473 (97)
2	2	2	2	.716 (74)	.870 (69)
1	1	1	2	.858 (366)	.841 (471)
1	1	2	2	.855 (441)	.817 (328)
1	2	1	2	.892 (102)	.617 (141)
1	2	2	2	.889 (261)	.836 (192)
2	1	1	1	.716 (1029)	.787 (875)
2	1	2	1	—	.824 (17)
2	2	1	1	.572 (159)	.523 (235)
2	2	2	1	—	.781 (14)

[a] Entries in parentheses are the frequencies on which each conditional is based.

from the marginal probabilities. The fact that the $P(A_1|S_iS_jA_2E_1)$ entries were, with one exception, the lowest in each group of four suggests a systematic rather than chance effect. Moreover, $\chi^2(11) = 23.59$ and $\chi^2(15) = 39.37$ for the π: 1 and .5 and .9 and .6 conditions.[2] The largest contributor to the former was $P(A_1|S_2S_2A_2E_1)$ with $\chi^2(1) = 9.18$, and $P(A_1|S_2S_1A_2E_1)$ contributed most heavily to the latter condition with $\chi^2 = 11.67$. Even though the cases contributing to each *df* may not have been mutually independent (i.e., each *S* could contribute to several categories) and tabled significance levels may therefore not apply, chi-squares of this magnitude would seem to require rejection of the independence hypothesis.

The relatively low values of $P(A_1|S_iS_jA_2E_1)$, taken together with the high values of $P(A_1|S_iS_jA_1E_2)$ and $P(A_1|S_iS_jA_2E_2)$, suggest the presence of stronger "lose-stay" and "win-shift" tendencies than are allowed by extant reinforcement models. It may be that the apparent absence of reinforcement effects in pooled conditional data from successive-cue situations actually represents a mixture of strategies adopted by individual *S*s rather than the operation of a single underlying rule for decision making.

REFERENCES

Atkinson, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–322.

Atkinson, R. C., Bogartz, W. H., & Turner, R. W. Supplementary report: Discrimination learning with probabilistic reinforcement schedules. *J. exp. Psychol.*, 1959, **57**, 349–350.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, 22, 133–145.

Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.

Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

Halpern, J. Stimulus similarity and reinforcement probability in three stimulus-two choice discrimination learning. Unpublished doctoral dissertation, University of Massachusetts, 1966.

Lee, W. Conditioning parameter model for reinforcement generalization in probabilistic discrimination learning. *J. math. Psychol.*, 1966, **3**, 184–196.

Lipkin, S. G. An analysis of the interaction of incentive and relative event frequency effects. Unpublished doctoral dissertation, University of Massachusetts, 1966.

Moore, J. W., & Halpern, J. Three stimulus-two choice auditory discrimination learning with blank trials. *J. exp. Psychol.*, 1967, **73**, 241–246.

Myers, J. L., & Atkinson, R. C. Choice behavior and reward structure. *J. math. Psychol.*, 1965, **1**, 170–203.

Popper, J., & Atkinson, R. C. Discrimination learning in a verbal conditioning situation. *J. exp. Psychol.*, 1958, **56**, 21–25.

Uhl, C. N. Effect of overlapping cues upon discrimination learning. *J. exp. Psychol.*, 1964, **67**, 91–97.

Wollen, K. A. Relationship between choice time and frequency during discrimination training and generalization tests. *J. exp. Psychol.*, 1963, **66**, 474–484.

(Received June 16, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 187–192

RECALL AND RECOGNITION IN LIST-DISCRIMINATION TASKS AS A FUNCTION OF THE NUMBER OF ALTERNATIVES[1]

NORMAN J. SLAMECKA

University of Vermont

In 2-list discrimination experiments, retention of 20-item lists was tested at 3 levels of item-population size (stimulus uncertainty), with 3 test procedures. Group 1 had free recall, Group 2 had full recognition (all alternatives present), and Group 3 had short recognition (40 alternatives present). Also, the effect of successive tests on the same material was examined. Results showed no difference between recall and full recognition performances. Short recognition had higher scores at high uncertainty levels. Retention was strongly influenced by stimulus uncertainty in all cases. Recall was not facilitated following either recognition procedure. Although results suggested an effect of response uncertainty, it was concluded that it may have been an artifactual consequence of differential chance benefits due to partial learning.

When memory for a list is tested by having *S* choose, from a larger set, only those items which were in the list, the performance affords a measure of discrimination of list membership. In effect, the larger set is to be subdivided into items from the list, and all others. The recognition procedure of testing retention is, in this sense, purely a test of list discrimination (Postman & Rau, 1957).

The procedure of free recall, on the other hand, imposes an additional requirement. Not only must correct items be discriminated from others which come to mind, but they must also be *available* on an unaided basis. That is, they must exist in *S*'s repertoire as integrated units, able to be produced, so that they can be discriminated from incorrect ones. Such unaided availability is not required for recognition tasks since the items are available on the recognition sheet. As Deese said: "Thus, recall is production plus some discrimination, while recognition is only discrimination [1963, p. 25]."

If a recall task is structured so as to employ only integrated and available items, then it should be functionally equivalent to a recognition task. However, a recognition sheet not only assures item availability, it also automatically restricts the number of alternatives from which choices are to be made (Davis, Sutherland, & Judd, 1961). By drawing items from populations whose limits are known to *S* (i.e., 2-digit no.), these characteristics of availability and restriction of alternatives can also be met within a recall framework, and the task then reduces fully to its recognition counterpart. In such a case there should be no difference between recall and recognition performance, since both procedures would involve comparable list-discrimination tasks. With such a rationale, and under such conditions, Davis et al. (1961) reported no superiority of recognition over recall in terms of *information* transmitted. However, the comparison was made at only one level of stimulus uncertainty, and one of the assumptions underlying their information measure has recently been questioned (Field & Lachman, 1966).

---

[1] This research was supported by National Science Foundation Grant GB-2590. The author is grateful for the assistance rendered by Edwin Coppage and the consultation by David Hill of the Mathematics Department.

The present experiments examined recall vs. recognition performance in list-discrimination tasks as a systematic function of the number of stimulus alternatives. On the input side, stimulus uncertainty was governed by the size of the item population from which a list was drawn. The larger that population, the greater the uncertainty about which items would be in the list. On the output side, response uncertainty was governed by the number of recognition alternatives on the test sheet. Again, the larger that number, the greater the uncertainty. Reductions in response over stimulus uncertainty were achieved by using recognition sheets with less than the maximum possible number of alternatives. Thus, response uncertainty was equal to or less than, but not greater than, the corresponding stimulus uncertainty.

Several perceptual identification and retention studies have shown effects of stimulus and response uncertainty (Garner, 1962), but there is little systematic comparison of verbal recall and recognition procedures within such a context.

The present experiments also examined a finding by Postman, Jenkins, and Postman (1948), who used items from relatively indeterminate and unintegrated sets (CVC trigrams). They reported that recall improved after a recognition test on the same material, and hypothesized that the recognition test acted as another learning trial, raising the strength of some items to the level needed for recall. From the present viewpoint this effect would be attributed jointly to a rise in item availability and to a drop in the number of functional alternatives, as a result of exposure to the recognition sheet. With the present materials, such an effect would be expected only when recognition involved less response uncertainty than recall, and only because it might reduce uncertainty for the recall test.

METHOD

Experiment I

Subjects.—Sixty-three students of psychology at the University of Vermont served as *S*s. They were tested in nine groups of seven each.

Materials.—Lists of 20 items were drawn randomly from each of three populations. Lists consisted either of states, 2-digit no., or letter-no. pairs formed by pairing every letter with a digit from 0 to 9. The respective sizes of these populations were therefore 50, 90, and 260. Each list was made into a 20-card deck, one item per card. For adequate sampling, three lists were drawn from each population and used equally across test conditions. Two types of recognition sheets were used. The full-recognition sheet had the total population of alternatives, arranged alphabetically for states, by size for 2-digit no., and alphabetically by size for letter-no. pairs. With this sheet, response uncertainty was equal to stimulus uncertainty. The short-recognition sheet contained 40 random alternatives from each population, arranged as above. Response uncertainty was thus constant across populations, and always less than the corresponding stimulus uncertainty. For recall, blank sheets were provided.

Procedure.—Group 1 had recall, Group 2 full recognition, and Group 3 short recognition. Every group had one list from each population, and only one type of initial retention test. Counterbalancing for list sequence required three groups for each test procedure. A total of 21 *S*s contributed to each of the nine conditions of the design. Immediately after every initial retention test was a second test on the same material. Recall was followed by full recognition, and both types of recognition were followed by recall.

After all *S*s were seated, each was given a prearranged pile of six sheets, face down. It was explained that a random list of 20 states (for instance) would be read aloud twice in succession, in a different order each time. It was requested that after the second reading the top sheet be turned up and *S* should write, in any order, all of the items. Twenty different items had to be written, even if some were guesses. This was done to equalize the tendency to guess under all conditions, for all *S*s. For the recognition procedures *S*s were to encircle the correct items, and to encircle exactly 20. When finished, *S* put the sheet at the bottom of the pile. The list was read at about a 2-sec. rate, the deck was shuffled, and read again. The next step was introduced only after all *S*s had finished. Then, the next sheet was turned up and retention of the same list was tested by the other procedure. This sequence was followed for all three lists in a session.

Experiment II

As will be seen, Exp. I showed large retention differences, apparently as a function of population size (stimulus uncertainty). However, this interpretation was tentative because of a perfect confounding of population size with class of item. Retention of states may have been superior to letter-no. pairs because of lower uncertainty, or because of some intrinsic internal structure (Garner, 1962). Resolution of this ambiguity required another experiment, one which used the *same* class of materials throughout while allowing variation of population sizes.

Subjects.—Seventy-two students of psychology at the University of Vermont served as *S*s. They were tested in randomly constituted groups ranging in size from 7 to 10 (due to appointment lapses). A total of 24 *S*s contributed to each of the nine conditions of the design.

Materials.—Lists were composed of 20 3-digit no. randomly drawn from population sizes of 40, 100, and 300 consecutive no., whose respective limits were 580–619, 150–249, and 300–499. Three lists were drawn from each population and used equally across test conditions. Full- and short-recognition sheets were prepared, the short one always with 40 random alternatives. Blank sheets were used for recall.

Procedure.—The procedure was the same as that for Exp. I. Every group had one list from each population, and only one type of initial retention test. Group 1 had recall, Group 2 full recognition, and Group 3 short recognition. List sequence was counterbalanced across conditions. As before, initial recall was followed by full recognition, and both types of initial recognition were followed immediately by recall. Prior to the two readings of a list, the population from which it came was described and the population limits written on a blackboard and left in full view. Twenty responses were required on all tests.

RESULTS

Since all *S*s had to give 20 responses per test, raw scores reflected both actual retention as well as chance successes. A correction formula was developed and applied to each raw score in order to estimate the true retention score. On the assumption that *S* remembered some items and guessed at the rest, it can be shown that the usual correction, $R - \frac{W}{N-1}$ (or in proportions, $1 - \frac{1-R}{1-P}$), underestimates retention in this type of situation, particularly in the midrange of scores. The new correction was based on a maximum likelihood estimator whose mathematical derivation is shown in the Appendix.

Experiment I

The left half of Fig. 1 shows the corrected initial retention performance of each group at each level of population size. Analysis of variance indicated no effect of test procedures, $F\ (2, 60) = 1.34$, $p > .20$; a very significant effect of population size, $F\ (2, 120) = 114.67$,

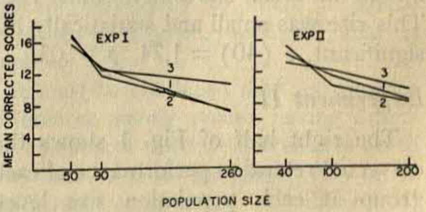


FIG. 1. Mean corrected initial retention performance as a function of population size (stimulus uncertainty). (Group 1 had free recall, Group 2 had full recognition, and Group 3, short recognition.)

TABLE 1

Mean Corrected Recall Performances in Both Experiments, Showing Scores when Recall was First, and when it Followed Full– and Short–Recognition Tests

Population Size	Recall First		Recall Second			
			Full		Short	
	M	*SD*	*M*	*SD*	*M*	*SD*
Exp. I						
50	15.75	2.17	13.72	3.72	14.22	3.86
90	12.78	2.57	11.07	3.89	10.58	4.06
260	7.04	3.92	6.97	4.24	8.05	3.40
Exp. II						
40	14.21	3.61	14.19	4.03	11.62	5.23
100	11.46	3.75	9.07	3.64	10.37	4.41
200	8.40	2.01	7.44	3.94	7.07	3.52

$p < .001$; and a significant Procedures × Population Size interaction, F (4, 120) = 2.71, $p < .05$. This interaction reflects the lesser slope shown by Group 3, suggesting a more consistent performance attributable to constant response uncertainty across population levels. However, an independent analysis of Group 3 alone nevertheless showed a highly significant influence of population size, F (2, 40) = 15.73, $p < .001$.

The pertinent data from the successive retention tests are seen in the upper half of Table 1 in mean corrected scores. Of the six possible comparisons, recall was lower after recognition in five cases. Only once, at the 260 population level following short recognition, did recall show an absolute rise. This rise was small and statistically insignificant, t (40) = 1.74, $p > .05$.

Experiment II

The right half of Fig. 1 shows the corrected retention performance of each group at each population size level. Analysis of variance again showed no effect of test procedures, F (2, 69) < 1; a highly significant effect of population size, F (2, 138) = 71.88, $p < .001$, which resolved the ambiguity concerning this same outcome in Exp. I; and a significant Procedures × Population size interaction, F (4, 138) = 4.30, $p < .005$. The interaction is again consistent with a flatter slope in Group 3, attributable to constant response uncertainty. A separate analysis of Group 3 alone still revealed a very significant effect of population size, F (2, 46) = 8.70, $p < .001$. Thus, the results of Exp. II confirmed in every way the findings of Exp. I.

Data from successive retention tests appear in the lower half of Table 1. Of the six possible comparisons, recall was lower after recognition in every case. These comparisons in both experiments agreed in showing no systematic tendency for recall to benefit from a preceding recognition test. Rather, a decline was the rule.

Discussion

The hypothesis that list-discrimination performances measured by recall and by recognition would be similar when the respective uncertainties were similar, was clearly supported by the equivalent scores shown in recall and full recognition across all levels of population size, in both experiments. These results reinforce the general position taken by Davis et al. (1961), with respect to the variables underlying that equivalence. Viewed solely as *procedures,* recall and recognition differ in only a formal, operational sense. What really matters for performance is the underlying task requirements which are implicitly manipulated along with these procedures. Inequalities in these attendant variables have probably been responsible for previously obtained performance differences between recall and recognition procedures. Thus, studies using poorly integrated items, such as CVC trigrams, have characteristically found recognition superior to recall (Luh, 1922; Postman et al., 1948; Postman & Rau, 1957). Such superiority can, in part, be attributed to the dependence of recall upon available, integrated items. Even with availability equated by using lists of common words, there is still a superiority of recognition over recall (Achilles, 1920; Postman & Rau, 1957). This may be attributed to the likelihood that more alternatives are entertained under recall than under recognition (no recognition sheet has ever included *all* of the common words). However, when both the availability and uncertainty are equated, as with Groups 1 and 2 above, then the superiority of recognition over recall disappears. It can be said that these testing procedures are themselves psychologically neutral, and serve only as vehicles for the functionally effective variables.

Consistent with this view, at least for list discrimination tasks, is the fact that recall did not improve after a recognition test. The hypothesis of Postman et al. (1948), that a recognition test affords rehearsal which benefits a subsequent recall test, was not confirmed under the present conditions. With alternatives equally numerous and available, as in full recognition and recall, there is no reason to expect any subsequent improvement in recall. Even after short recognition, where an effect might be expected through some reduction in response uncertainty, recall was not enhanced. Evidently the effect depends more upon increases in item availability rather than upon reduction of uncertainty.

Perhaps the most striking aspect of these data is the decisive control exerted by the size of the population from which a list was drawn (stimulus uncertainty). The influence of this variable was considerable, since in both experiments Groups 1 and 2 retained only about half as much under high uncertainty as under low uncertainty. Group 3 showed less change because of the constant response uncertainty, but it still showed significant variation due to stimulus uncertainty. The psychological mechanism which mediated these effects was not revealed, but possibly it was in the input encoding stage. The efficiency of encoding may vary with the size of the background set. Thus, if input were encoded by noting the positions within the set that list items occupied, this might be accomplished more precisely with a smaller set.

One final observation concerns recognition as a function of the number of alternatives on the test sheet. The flatter slope of Group 3, as compared to Group 2, is consistent with previous reports that recognition is inversely related to the number of response alternatives (Murdock, 1963; Postman, 1950; Teghtsoonian, 1965), even after correction for guessing. However, higher scores are necessarily to be expected when there are fewer alternatives, simply because of the increased chance probability that *partially* learned items will be correctly chosen (McNulty, 1965). Neither the usual nor the present corrections for guessing can compensate for this *differential* chance advantage due to partial learning, since they both assume that items are either known completely or not at all (Brown, 1965). This fact does not, however, invalidate the comparisons between recall and full recognition groups, since they were equated on number of alternatives. Murdock's (1963) hypothesis that *S* eliminates some items as incorrect and then chooses randomly from the remainder, also predicts better performance with fewer alternatives, even after a guessing correction. But how can *S* successfully eliminate some items as incorrect? Patently, by having at least partial knowledge of the correct items. Therefore, Murdock's hypothesis appears necessarily to entail a partial learning assumption, with all that it implies.

It must be concluded that performance differences among groups having different numbers of response alternatives (even after correction for guessing), may be an artifactual consequence of differential chance inflation of scores due to partial learning. At present there is no procedure which can estimate actual retention levels in such situations.

APPENDIX

Let N = total no. of alternatives, C = no. of choices to be made, Y = no. of correct choices, and X = no. known by S. The problem is: Find X, $O \leq X \leq Y$ such that $P[Y/X]$ is a maximum. This value is the maximum likelihood estimator of the true parameter value. The procedure is: Find the largest integer value of X for which $\frac{P[Y/X+1]}{P[Y/X]} > 1$, and denote this value by $\hat{X}$.

The desired estimator is the larger of the two no. $(O, \hat{X}+1)$. In general,

$$P[Y/X] = \frac{\binom{C-X}{Y-X}\binom{N-C}{C-Y}}{\binom{N-X}{C-X}}$$

$$\therefore \quad \frac{P[Y/X+1]}{P[Y/X]} = \frac{\dfrac{\binom{C-(X+1)}{Y-(X+1)}\binom{N-C}{C-Y}}{\binom{N-(X+1)}{C-(X+1)}}}{\dfrac{\binom{C-X}{Y-X}\binom{N-C}{C-Y}}{\binom{N-X}{C-X}}} = \frac{(Y-X)(N-X)}{(C-X)^2} \geq 1$$

implies $NY - NX - XY \geq C^2 - 2XC$, or finally, $\frac{(NY - C^2)}{(N+Y-2C)} > X$. The maximum likelihood estimator of X is the larger of $\left(O, \left[\frac{NY-C^2}{N+Y-2C}\right]+1\right)$.

REFERENCES

Achilles, E. M. Experimental studies in recall and recognition. *Arch. Psychol.*, 1920, **6**, No. 44.

Brown, J. A comparison of recognition and recall by a multiple-response method. *J. verbal Learn. verbal Behav.*, 1965, **4**, 401–408.

Davis, R., Sutherland, M. S., & Judd, B. R. Information content in recognition and recall. *J. exp. Psychol.*, 1961, **61**, 422–429.

Deese, J. Comments on Professor Murdock's paper. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 22–25.

Field, W. H., & Lachman, R. Information transmission (I) in recognition and recall as a function of alternatives (k). *J. exp. Psychol.*, 1966, **72**, 785–791.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31** (3, Whole No. 142).

McNulty, J. A. An analysis of recall and recognition processes in verbal learning. *J. verbal. Learn. verbal Behav.*, 1965, **4**, 430–436.

Murdock, B. B., Jr. An analysis of the recognition process. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 10–22.

Postman, L. Choice behavior and the process of recognition. *Amer. J. Psychol.*, 1950, **63**, 576–583.

Postman, L., Jenkins, W. O., & Postman, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, **61**, 511–519.

Postman, L., & Rau, L. Retention as a function of the method of measurement. *Univer. Calif. Publs. Psychol.*, 1957, **8**, 217–270.

Teghtsoonian, R. The influence of number of alternatives on learning and performance in a recognition task. *Canad. J. Psychol.*, 1965, **19**, 31–41.

(Received June 20, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 193–198

WORD FREQUENCY AND ACCUMULATIVE PROACTIVE INHIBITION

BENTON J. UNDERWOOD AND BRUCE R. EKSTRAND[1]

Northwestern University[2]

An interference theory of forgetting which emphasizes PI from linguistic associations predicts that the PI in the recall of successive lists of words should build up more rapidly if the lists are made up of high-frequency words than if made up of low-frequency words. Ss learned 4 lists of 16 pairs to a criterion of 12 correct, each successive list being recalled after 24 hr. The words in the lists learned by 1 group were AA 2-syllable words; for the other group, the frequency was 1 per million. PI increased directly as a function of number of prior lists but did not differ as a function of word frequency; thus, theoretical expectations were not confirmed.

This study was predicated on the assumption that proactive inhibition (PI) in the recall of word lists increases as a function of the number of prior lists learned in the laboratory. The rate of accumulation of PI might be expected to vary depending upon the associative relationships which obtain among the successive lists. An investigator has two general alternatives in studying the effect of these associative relationships. First, he may use particular paradigms for which the associative relationships for successive lists can be precisely specified. Thus, the recall of each successive list could be determined when the lists conform to an A-B, A-C paradigm, with the recall being compared, perhaps, with that obtained when the lists fit an A-B, C-D paradigm. The second alternative is to study the effect of associative relationships which have developed among words through normal linguistic usage. Although the first approach has a distinct analytical advantage over the second, the second has the advantage of representativeness—representativeness of retention functions as they might be expected to occur outside the laboratory. Because of theoretical considerations, the second approach was chosen for the present experiment.

One version of an interference theory of forgetting assumes that PI can be produced by conflicting linguistic associations which *S* has acquired outside the laboratory (Underwood & Postman, 1960). Tests of this theory have been based upon the belief that lists can be constructed which will be interfered with in varying degrees by preexperimental linguistic associations. The results of most such tests have failed to give substantial support to the theoretical expectations (Ekstrand & Underwood, 1965; Postman, 1962; Underwood & Keppel, 1963), although there are exceptions (Turnage & Anderson, 1965). There are a number of possible reasons why the theory may be correct in principle while remaining unconfirmed by experiment. One such reason is that laboratory-acquired associations may, in some manner, remain insulated from potentially interfering associations acquired outside the laboratory. The obvious test of such a notion is to bring into the laboratory the linguistic associations which are presumed to produce interference. This is the approach chosen for the present study.

---

[1] Now at the University of Colorado.

[2] This work was done under Contract Nonr-1228(15), Project NR 154-057, between Northwestern University and the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

The number of associations among a sample of high-frequency (HF) words is known to be greater than the number among a sample of low-frequency (LF) words (e.g., Deese, 1960). Because of the greater number of interitem associations among HF words than among LF, the probability of the associations producing interlist interference at recall should be higher for lists of HF words than for lists of LF words. If, then, *S* learns and recalls each of a series of lists made of ostensibly unrelated HF words, the build up of PI should occur more rapidly than for a series of lists made from LF words.

METHOD

Each *S* served 5 successive days, Monday through Friday. On Monday, the first paired-associate list of 16 pairs was learned to a criterion of 12 correct responses. On Tuesday, 24 hr. later, *S* recalled the list learned the previous day, then immediately learned a second list. This second list was recalled on Wednesday, a third list was learned, and so on. One group of *S*s was assigned to lists made up of HF words, another to lists made up of LF words.

Lists.—Eight lists of 16 pairs each were constructed, four from two-syllable AA words and four from two-syllable words having an occurrence of once per million (Thorndike & Lorge, 1944). Pools of words at each frequency level were developed by taking the first two-syllable word appearing on each successive page. Proper names and words with two different spellings were excluded. From each pool 128 words were drawn, using a table of random numbers, with each successively drawn two items being paired. The first 16 pairs became one list, the next 16 a second list, and so on.

Procedure.—The study-recall method was used in learning. On study trials each pair was presented for 2 sec., and on recall trials, alternating with study trials, the stimulus terms were presented for 2 sec. with *S* instructed to produce the appropriate response term. The learning of each list was carried until 12 correct responses were given on a recall trial. Three different orders of the 16 pairs were used on study trials and three different orders of the stimuli on recall trials.

Each successive list was given a single recall trial 24 hr. after it had been learned. The *S* did not see the response terms on this recall trial, and no relearning was given. Following the recall of the fourth list, additional recall tests were undertaken. First, *S* was given a sheet with 16 spaces available for each of the four lists, identified by number in order of learning, and was asked to write down all of the pairs possible under the appropriate list number. On a further test, *S* was given a new sheet with all 64 stimuli listed in random order with a blank after each. The *S* was first asked to write down all responses possible after the appropriate stimulus terms. Following this, *S* was requested to identify each stimulus term by writing down the number of the list of which it was a member (first, second, third, fourth). On the third and final stage, a matching procedure was used. The 16 responses appropriate for a list were listed along with a list of the 16 stimuli. Each response term was numbered and *S*s' task was to place the appropriate number after each stimulus term. The matching task was independent for each list although all items for the four lists were given on a single sheet and *S* could work on the lists in any order he chose.

Design.—Two groups of 24 *S*s each were used, one group assigned to the HF lists, the other to the LF, assignment being made on a random basis. The four lists were completely counterbalanced. All *S*s had had previous experience in verbal-learning experiments.

RESULTS

Learning.—The mean numbers of trials to learn the four successive lists at each frequency level are shown in Table 1. It should be clear that list numbers represent the temporal order only, since lists were counterbalanced. There is no significant source of variance among the eight means. The *F*s for both frequency and stage of practice (lists) were less than one. The interaction (1.71) was far short of the 5% significance level. Thus, it is con-

TABLE 1

MEAN TRIALS TO LEARN AS A FUNCTION OF FREQUENCY AND STAGE OF PRACTICE

	List 1		List 2		List 3		List 4	
	M	σ	M	σ	M	σ	M	σ
LF	6.83	3.94	6.17	2.86	6.92	5.98	6.46	3.24
HF	5.04	3.86	6.46	4.52	5.54	2.44	5.92	3.23

cluded that word frequency is not a relevant variable in learning these paired-associate lists and there is no evidence of learning to learn.

The mean numbers of overt errors per trial did not differ as a function of frequency. The only significant source of variance was stage of practice, F (3, 138) = 4.40, $p < .01$. For both frequency levels, more errors were made on the first list than on the remaining three, with no consistent differences among the latter. The only appreciable difference in types of errors consisted of giving a stimulus term as a response. The total such errors for the HF lists was 122, for the LF lists, 44. These represent 24% and 10% of the total errors, respectively.

Paced recall.—The fact that the trials to learn did not differ as a function of either frequency or stage of practice, plus the use of the study-recall method to minimize learning on the criterial trial, makes the use of recall scores appropriate as indexes of forgetting. However, because there were slight differences in mean performances on the criterial trial, loss scores are used to measure forgetting, these being the difference between the number correct on the criterial trial and the number correct at recall. The mean loss scores are shown in Fig. 1. It is apparent that forgetting increases as a function of stage of practice—PI increases as a direct function of the number of prior lists. However, it is further apparent that there are no consistent differences between the HF and LF lists. Analysis of variance shows only one F above one, namely, that for stage of practice. This F (3, 138) = 12.95 is significant far beyond the 1% level. Both groups show an increase of approximately 20% in forgetting between the first and fourth list.

Two facts are worth noting concerning the overt errors at recall. For all four lists combined, 23 stimuli from the list being recalled were given as responses by *S*s having the HF lists, only one by *S*s having the LF lists. A total of 33 intrusions from previously learned lists occurred in the recall of the HF lists, 16 in the recall of the LF lists. The groups did not differ appreciably in misplaced response errors.

Unpaced recall.—The unpaced-recall tests were administered following the

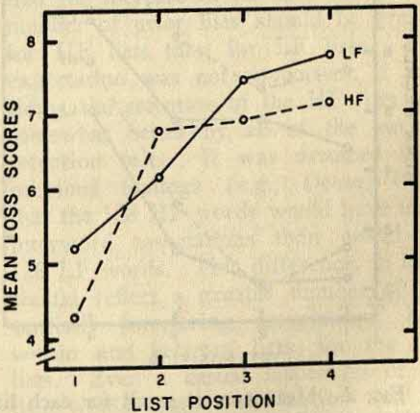


FIG. 1. Loss scores over 24 hr. as a function of list position and word frequency on paced recall.

single-paced recall test on the fourth list. In the first phase of unpaced recall *S* was asked to write down all of the pairs he could remember under the appropriate list number, but was also instructed to write down single words even if he did not know the list to which they belonged or the words with which they were paired. The total number of words recalled was far greater for the HF lists than for the LF lists, the means being 37.21 and 23.67, respectively. The total number of paired words produced was also greater for the HF lists, as was also the number of correct pairings, these latter means being 13.54 and 7.16. However, a calculation of the number of correct pairs per number of pairs produced results in only a small and insignificant difference as a function of frequency, the mean values being 78.7% for HF, 74.2% for LF. Finally, the mean numbers of correct pairs which were also correctly identified as to list were determined. For the four lists in order, the mean values were .67, .58, .92, and 3.75 for the LF lists, and 1.29, 1.33, 2.21, and 5.42 for the HF lists. Both lists ($F = 60.15$) and frequency ($F = 10.91$) are significant statistically. Recall from the first three lists is, in an absolute sense, very low at both frequency levels. The picture that emerges from this first unpaced recall is a greater availability of words and pairs from the HF lists than from the LF lists. However, given that a pair is recalled, its likelihood of being a correct pairing does not differ as a function of frequency.

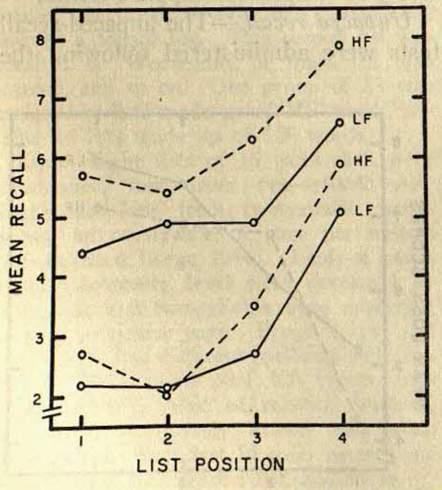


Fig. 2. Mean correct recall for each list when stimuli are provided on an unpaced test. (The upper two curves disregard list designations; the lower two curves result when correct list designation is required.)

In the second unpaced recall, stimuli in all four lists were listed randomly and *S* was asked to supply as many correct responses as possible and to designate the list number for each stimulus term whether the response was recalled or not. As might be expected, providing *S* the stimulus term both increases recall and reduces the difference between LF and HF. The basic findings are shown in Fig. 2. The upper two curves give the mean number correct disregarding errors in list designation; the lower two curves result when correct list designation is added as a further criterion of correctness. The difference between the upper two curves is significant statistically, $F\ (1, 46) = 4.76$, $p < .05$, that between the lower two curves is not ($F = 1.69$). List differences are significant in both cases, interactions are not. The values for the recall of the fourth list in the lower two curves are almost exactly the same values as observed on the paced-recall test for these lists.

Further analyses were undertaken to determine what variables were involved in supplying correct-list designations. If the correct response is produced to the stimulus term, is the list designation more likely to be correct than if an incorrect (or no response) is given? The answer is "yes." Correct-list designations were supplied to 54.6%

of the HF items when the correct response was present, but for only 33.5% when it was not. The values for the LF lists were almost identical, being 57.0% and 35.5%. Clearly, the presence of the response term is associated with *S*s' ability to correctly identify list membership.

Although the data will not be presented, it can be reported that the probability of producing a correct response is directly related to the number of times it was given correctly in learning. The further question asked about list identification is: "Given a correct response to a stimulus term on the unpaced test, is the number of times the response was given correctly during learning related to correct-list identification?" The answer to this question is "no." The pairs for each *S* were sorted into one of four categories based upon the number of times given correctly during learning. Thus, there were four "strength" categories, each consisting of four items for each *S*. A determination was then made for each category of the percentage of correct responses on the unpaced recall which were also correctly identified as to list. This tally was made for the first three lists combined. For the four categories, from weakest to strongest, the percentages reflecting correct list identification were 43, 50, 48, and 44 for the HF lists; 50, 42, 46, and 51 for the LF lists. Thus, while the number of correct responses produced is directly related to the number of times a response has been given correctly in learning, the proportion of those correctly identified as to list is not.

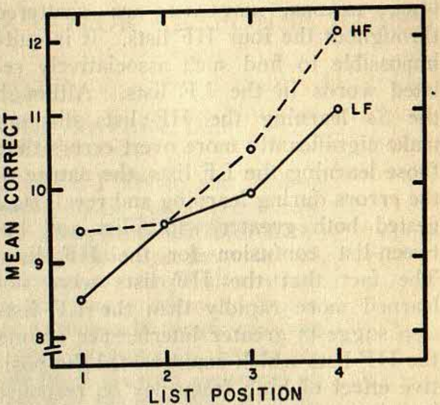


Fig. 3. Mean number of correct matchings as a function of list position and word frequency.

Associative matching.—The final retention data came from a matching procedure wherein *S* was given the stimulus and response terms for each list, and was asked to pair them correctly. This procedure neutralizes any differences in response availability as a function of frequency and measures in more or less pure form the integrity of the associations between stimulus and response terms. The findings are given in Fig. 3. Although the number of correct matchings is somewhat greater for the HF lists than for the LF lists, the *F* (1.10) is far from significant. Only list position produces a significant effect ($F = 8.28$).

Discussion

The theoretical considerations outlined in the introduction led to the expectation that the increase in PI as a function of number of prior lists should be greater for HF lists than for LF lists. This expectation was not supported; if anything, the retention of the HF lists was somewhat better by all of the various retention tests. It was assumed from previous findings (e.g., Deese, 1960) that the 128 HF words would have more interword associations than would the 128 LF words. This difference, in turn, should reflect a greater number of potentially interfering associations, both within and between lists, for the HF lists. Even a casual inspection of the lists would support the notion that the HF words have more interword associations than the LF words. For example, the words BATTLE, SERVICE, ARMY, SOL-

DIER, LEADER, CAPTAIN are scattered throughout the four HF lists. It is quite impossible to find such associatively related words in the LF lists. Although the *S*s learning the HF lists did not make significantly more overt errors than those learning the LF lists, the nature of the errors during learning and recall suggested both greater within-list and between-list confusion for the HF lists. The fact that the HF lists were not learned more rapidly than the LF lists also suggests greater interference among the HF lists which counteracted the positive effect of high frequency on response learning as such. And clearly, the unpaced-recall tests showed availability of both stimulus terms and response terms was greater for the HF lists. Thus, the number of potentially interfering items would seem to increase across lists more rapidly for the HF lists than for the LF lists. Yet, the fact remains that the retention results do not show that PI built up more rapidly for the HF lists.

The possibility remains, of course, that there are sources of interference in the LF lists which are not present in the HF lists, and that these sources are as potent as those in the HF lists. For example, one might judge the formal similarity to be greater among the words in the LF lists than among those in the HF lists. If this is true, it becomes apparent that an interference theory based only on the role of naturally developed linguistic associations among words and among letters is incomplete, and that other potential sources of interference which vary as a function of word frequency (and perhaps other characteristics) must be identified and subsequently given analytical study. In short, as have the data from other studies noted in the introduction, those from the present one make obvious the need for modifying assumptions about interference based on linguistic habits.

Finally, certain facts about PI as such must be reviewed. It has been suggested (e.g., Warr, 1964) that evidence for a build up of PI with successive lists (ostensibly unrelated) is unconvincing. It is said to be unconvincing because the degree of learning on successive lists taken to the same criterion may differ as a consequence of learning to learn. The argument is that the more rapid the learning to a given criterion the less the degree of learning. While the opposite relationship is more likely correct, the matter can be resolved directly. Both in an earlier study (Underwood & Schulz, 1961), and in the present one, the number of trials to attain a common criterion did not differ as a function of stage of practice and still PI increased directly as a function of number of prior lists learned.

REFERENCES

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.*, 1960, **7**, 337–344.

EKSTRAND, B. R., & UNDERWOOD, B. J. Free learning and recall as a function of unit-sequence and letter-sequence interference. *J. verbal Learn. verbal Behav.*, 1965, **4**, 390–396.

POSTMAN, L. The effect of language habits on the acquisition and retention of verbal associations. *J. exp. Psychol.*, 1962, **64**, 7–19.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

TURNAGE, T. W., & ANDERSON, N. S. Letter-frequency and associative probability as determinants of learning and retention. *J. verbal Learn. verbal Behav.*, 1965, **4**, 463–468.

UNDERWOOD, B. J., & KEPPEL, G. Retention as a function of degree of learning and letter-sequence interference. *Psychol. Monogr.*, 1963, **77**(4, Whole No. 567).

UNDERWOOD, B. J., & POSTMAN, L. Extraexperimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235.

WARR, P. B. The relative importance of proactive inhibition and degree of learning in retention of paired associate items. *Brit. J. Psychol.*, 1964, **55**, 19–30.

(Received June 20, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 199–202

SEQUENTIAL COMPLEXITY AND MOTOR RESPONSE RATES

FRED L. ROYER

Veterans Administration Hospital, Brecksville, Ohio

Keeping pace with an auditory click presented 1 per sec. and accelerated at a rate of .2 per sec. per 8 clicks, 16 Ss reproduced binary sequences, by pressing 2 telegraph keys. Mean maximum rates of response varied 3.32–5.61 per sec. The explanation of the order of difficulty holds that runs of the same element and single alternations constitute response units (RUs). Response rate is influenced by the number of junctures of RUs in a sequence: those with longer runs are easier and have few junctures. At a more analytic level, longer sequences are strings of RUs drawn from the 4 different 2-element or the 8 different 3-element sets. Mixtures of 2- and 3-element RUs increase the difficulty. Where the number of junctures in a sequence is the same, the sequence having repetitions of the same RU is easier.

Royer and Garner (1966) presented *S*s with auditory eight-unit binary sequences repeated continuously. When *S*s felt able to reproduce the sequence, they pressed two keys in synchrony with the stimuli. The perceptual difficulty of a sequence, measured by the delay before responding and errors after responding, was related to the number of alternative psychological organizations or reorganizations of the sequences into patterns, as measured by response uncertainty (the information theoretic measure of the variability of the frequency of use of each unit in the sequence as a point at which to begin responding). That perceptual difficulty was being measured, not the difficulty of producing the responses, is indicated by the finding that some *S*s executed nearly all patterns without errors; yet the relative delays before responding to various patterns by these *S*s were proportional to the relative delays by *S*s making false starts and asynchronous errors.

Sequences of motor responses obviously do vary in difficulty. It is known from time-motion studies, e.g., that certain sequences facilitate the speed of execution of a task. In the performance of music some melodic and rhythmic sequences are harder to perform than others. In these tasks the difficulty of performance is related to the maximum speed of executing the motor sequence.

The purpose of this study was to investigate the difficulty of performing a series of repeated motoric sequences as a function of the complexity of the sequences. Difficulty is defined by the maximum rate of responding correctly; complexity is defined by results of the study by Royer and Garner (1966). The use of the results of the study of the perceptual organization of auditory sequences to define the complexity of a sequence for a study of motor performance is predicated upon the assumption that sequences, regardless of their source, contain information which is processed similarly by the central nervous system.

METHOD

Subjects.—Sixteen students of a course in general psychology at Kent State University served as *S*s in this study.[1] Seven *S*s were

---

[1] The author wishes to express his gratitude to Joseph Grosslight, Chairman, Department of Psychology, Kent State University, for his kind assistance in providing *S*s and space.

men and 9 were women. All were right handed. A total of 18 *S*s were run, but the data from the first 2 *S*s could not be used because of procedural problems.

Sequences.—All 20 of the Royer-Garner patterns were used. The particular sequences drawn from these patterns were those which their *S*s used with the highest frequency. For example, the sequence drawn from Pattern K (11110010) was Sequence K5 (00101111) since regardless of which of the eight sequences of Pattern K were presented, most *S*s chose K5 as their sequence of response. The sequences used are shown in Table 1. In addition, all eight different sequences forming Pattern T were used.

Apparatus.—The *S* responded by pressing two telegraph keys with his preferred hand. His rate of response was determined by auditory clicks. The clicks were generated by square-wave pulses, 10 msec. wide, from a Grass stimulator S4 coupled to an audio amplifier. The output of the stimulator was also used to drive a counting unit which indicated the occurrence of eight clicks by flashing a light. This light was used as a signal by *E* to advance the frequency of the pulse generator. Increments in rate were .2 pulses per sec., made manually after each eight clicks. The occurrence of an error by *S* was determined with a digital logic system. Using a system of comparators, *S*'s responses were matched against a setting of switches corresponding to the sequence presented to *S*. The matching of the response with the switch setting was accomplished by a ring counter which advanced with each key depression by *S*. Mismatch of switch setting and key response activated a light signaling an error. The click generator was then turned off.

TABLE 1

MAXIMUM RESPONSES PER SEC. FOR BINARY SEQUENCE

	Sequence	*M*	*SD*
A1[a]	11111111	5.61	0.81
A1′	00000000	5.54	0.95
F1	11110000	5.41	1.21
E1	11111110	5.24	1.22
H1	11111100	5.04	1.15
B1	10101010	4.90	0.77
G1	11111000	4.89	1.17
I4	01010111	4.79	1.38
C1	11001100	4.58	1.09
D1	11101110	4.56	1.23
L6	01011111	4.53	1.69
K5	00101111	4.46	1.19
M1	11110100	4.34	0.96
N5	01101111	4.29	1.67
P4	00110111	4.24	1.35
O1	11101100	4.04	1.14
J1	11101000	4.01	1.54
R3	01010011	3.94	1.59
S7	10110100	3.69	1.45
Q4	00100111	3.64	1.29
T1	11011010	3.32	1.06
T1	11011010	3.83	1.11
T2	10110101	3.16	1.32
T3	01101011	3.61	1.34
T4	11010110	3.57	1.20
T5	10101101	3.40	1.42
T6	01011011	3.74	1.34
T7	10110110	3.57	1.18
T8	01101101	3.37	1.16

[a] Identification of sequences after Royer and Garner (1966).

Because of the need to operate several independent circuits from the response key, the keys drove a relay. Unavoidably, this resulted in a differential sound when the keys were pressed, so that auditory information about *S*'s performance was available to both *S* and *E*.

Procedures.—The *S*s were told that the task they were about to do involved making a series of movements at a particular rate. They were told to press the two keys with one hand. (Most *S*s used the first three fingers to operate the keys; some used only the index finger for both keys.) One key was labeled "1," the other "0." The *S* was given a card displaying Sequence A, 11111111. He was told to press the key marked "1," to keep up with the speed of the clicker, and to continue without pause until told to stop. Beginning at one pulse per sec., the clicker was turned on and *S* began responding. After *S* made an error or reached the point where he was unable to maintain the same rate of response as the clicker, the trial was discontinued. The next sequence presented was the complement of A, (A′), 00000000. The procedure was repeated subsequently with Sequences B1, C1, and D1. The remaining Sequences, E1–T1, were presented to *S* in a randomized order. At the completion of the presentation of all these sequences, including T1, the eight sequences comprising Pattern T were presented in randomized order. The randomization was achieved by shuffling the cards. The *E* was able to determine the point at which *S* could not keep in synchrony with the clicker by listening to the sounds of the relays activated by the keys; this determination was necessary only with the several very simple sequences.

RESULTS

In Table 1, the *M*s and *SD*s of maximum response rates for each sequence are shown. They are presented in order of decreasing maximum rates, i.e., in order of increasing difficulty. The identification of the sequences is that used by Royer and Garner (1966).

The differences between all pairs of means excluding A and A′, were tested with Duncan's new multiple-range test ($s_{\bar{x}} = 0.16$, $k = 19$, $df = 270$, $\alpha = .05$), using the error term of an analysis of variance of repeated measures. The subsets of means within *nonsignificant* ranges are: FEH, EHBGI, HBGICDL, BGICDLK, ICDLKMN, CDLKMNP, LKMNPO, KMNPOJ, MNPOJR, OJRSQ, and SQT.

The data show that maximum rate is greatest when there are long repetitions of a single response. Sequence A1 is the first sequence presented and might be expected to lack a practice effect; however, it has the most rapid response rate. Next in order of difficulty are Sequences F1 (11110000), E1 (11111110), and H1 (11111100), which have long runs of a single element.

In the sequences which follow the above in order of difficulty, those containing strings of simple alternations B1 (10101010) and I4 (01010111), and those formed from repetitions of sequences of two- or four-unit lengths, C1 (11001100) and D1 (11101110), are interspersed among those having long runs of a single element, G1 (11111000) and L6 (01011111). This ordering of sequences suggests that long runs of the same element, strings of alternations, and symmetrical sequences facilitate performance.

Among those sequences which have an eight-unit length before repetition, those which begin with a number of short runs, such as I4, L6, N5, R3, and S7, appear to be associated with greater variability.

One of the main problems of interest was the degree of correlation which might be found between measures of difficulty of motoric responding and measures of perceptual difficulty. Product-moment correlations were computed for the maximum rates of responses and the summary measures of errors, delay, and response uncertainty obtained by Royer and Garner (1966). The correlations were computed for only those sequences eight-unit long before repetition (E-T). The mean maximum rate of response was correlated −.77 with errors, −.68 with delay, and −.91 with response uncertainty. The correlation with errors and delay indicated that perceptual difficulty and response difficulty of the sequence are highly related, affirming our expectations. The correlation of maximum rate of response with response uncertainty indicates that the complexity of the sequence has comparable effects on both perceptual and motor activities. This also confirms our expectations.

The appraisal of the effects of different starting positions of the Pattern T on response rate showed that starting responses at the beginning of a run or a single alternation facilitates responding, except when the sequence begins and ends with the same element, as does T2 (10110101).

DISCUSSION

In the Royer-Garner study, *S*s were free to begin responding at any of the eight different places in the sequence. The places with the highest frequency of use were always at the beginning of simple alternations or at the transition from runs of one stimulus element to runs of the other element. Similar results were found by Garner and Gottwald (in press) who, using a guessing-game technique similar to that used in probability learning, had *S*s learn sequences of binary events (lights) produced by repetition of patterns of five-unit length. They found that the sequences were organized into patterns beginning with a run, the longer run being preferred, or into patterns beginning with a pair of single alternations.

It appears that runs and simple alternations constitute perceptual units. A point in a sequence where a transition from one such unit to another occurs constitutes a juncture or point of articulation around which to organize the sequence into a pattern. As the number of junctures increases, the response uncertainty increases.

In this study, the maximum rate of response is highly correlated, inversely, with the response uncertainty for the patterns in the Royer-Garner study. As the number of junctures increases, the speed of performance decreases. Perhaps simple alternations and runs constitute not only perceptual units, but response units (RUs) as well.

Here, the data for Pattern T in all its starting points show that when a sequence begins with the longer run, as T1 and T4, or with a single alternation, as T6 and T7, the response rates are relatively fast and stable. Although T2 and T5 begin with a single alternation, these sequences are difficult. As such a sequence is continued the alternation becomes a run; in order to preserve the identity of the repeated sequence, *S* must make a "break" in a run. Making a break in a run would require forming new RUs, probably by the introduction of additional response variables such as timing or force.

Borrowing from rhythmic analysis, we may note that most if not all sequences can be described as being formed from the juxtaposition of two-unit (duple) or three-unit (triple) sequences. The total set of duple sequences is 11, 10, 01, and 00. The total set of triple sequences is 111, 110, 101, 100, 011, 010, 001, and 000. Thus, a pattern such as I4 may be thought of as being composed of the juxtaposition of two subsequences drawn from the set of duple sequences, i.e., 01 is repeated three times and followed by 11. The most difficult sequence, T1, may be thought of as being composed from the juxtaposition of sub-sequences drawn from the set of triple sequences, 110, repeated once and followed by one drawn from the set of duple sequences, 10. Pattern 04 is the next most difficult; it may be thought of as having a similar composition. S4 would be a sequence which used all four of the duple sequences. R3 is composed of three of the duple sequences.

There is a suggestion, in other words, that eight-unit binary sequences are difficult to execute when the sequences are composed of (*a*) a mixture of duple and triple sequences or (*b*) a string of unrepeated and different duple sequences.

It may be more appropriate to think of RUs as being duple and triple motor sequences. Long runs can just as readily be broken down into such units as strings of alternations or mixtures of both runs and alternations. Analysis in terms of such RUs suggests that the locations of errors should be predictable. The errors should occur with higher frequency at the junctures of unrepeated RUs and at junctures making a transition from a triple to a duple RU or vice versa. Also they should occur with higher frequency at the element of a RU which is different from the element in the same position in the preceding RU when that preceding RU has been repeated.

Unfortunately, these predictions can not be tested on the present data. The design and instrumentation of the experiment did not permit measurement of the location of errors because there would have been too few errors per sequence to allow an adequate statistical evaluation.

REFERENCES

GARNER, W. R., & GOTTWALD, R. L. Some factors in learning of sequential patterns of binary events. *J. verbal Learn. verbal Behav.*, in press.

ROYER, F. L., & GARNER, W. R. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Percept. Psychophys.*, 1966, **1**, 41–47.

(Received June 21, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 203–211

RETROACTIVE INHIBITION IN FREE RECALL[1]

LEO POSTMAN AND GEOFFREY KEPPEL
University of California, Berkeley

An investigation of retroactive inhibition (RI) in free recall is reported. The design comprised 3 experimental groups and a control group. The experimental groups learned 2 successive lists of 20 2-syllable nouns. The 1st list was presented for 4 alternate study and test trials, and the 2nd list for either 2, 4, or 6 such cycles. On the terminal test of retention Ss were required to recall both lists. The control group learned and recalled a single list. Recall for the 1st list decreased progressively as a function of the number of interpolated trials whereas recall for the 2nd list increased. There was also a progressive tendency for List-2 items to precede List-1 items in recall. Identification of list membership at the end of recall remained highly accurate throughout. Thus, interpolated learning produces effects in free recall which are functionally continuous with unlearning in paired-associate situations. RI in free recall is attributed to the loss of contextual and interitem associations.

This study investigates retroactive inhibition (RI) in the free recall of lists of words. The theoretical question of interest is whether and to what extent RI in free recall exhibits functional characteristics which parallel those observed in studies of unlearning with paired associates. Unlearning refers to the reduction in the availability of first-list items as a consequence of interpolated learning (IL) of a second list. Analysis of the conditions of unlearning (McGovern, 1964) has led to the conclusion that IL disrupts associations between first-list responses and (*a*) specific stimulus terms and (*b*) contextual or environmental stimuli. The distinction between the two classes of associations is supported by the variations in the amount of unlearning obtained as a function of the paradigm of transfer and the method of testing. On an unpaced test of recall retention losses are substantially greater for the A-B, A-C paradigm, in which both classes of associations are subject to unlearning, than for the A-B, C-D paradigm in which only contextual associations can be affected. When the first-list responses are supplied on a test of associative matching, retention losses are eliminated altogether for A-B, C-D, and greatly reduced for A-B, A-C. Within the framework of this analysis, unlearning in free recall would be attributed primarily to the loss of contextual associations. It is likely, however, that the unlearning of specific associations contributes to RI in free recall. If individual words become associated or are grouped into higher-order units (Tulving, 1962b), semantic or conceptual similarity between functional stimuli requiring different responses in the successive lists would create conditions of negative transfer for specific interitem associations. These functional stimuli might be mediators, such as the names of

---

[1] This research was supported by Contract Nonr 222 (90) between the Office of Naval Research and the University of California and a grant (MH-10249) from the National Institutes of Health. The work was carried out at the Institute of Human Learning which is supported by grants from the National Institutes of Health and the National Science Foundation.

categories, or words in the list which serve to elicit groups of related items.

When successive lists of paired associates conform to the A-B, A-C paradigm, unlearning is assessed by means of an unpaced test or recall (MMFR) in which *S* is presented with the common stimulus term (A) and required to recall both responses (B and C) in the order in which they occur to him (Barnes & Underwood, 1959). To obtain an optimal estimate of the availability of first-list responses in the experimental situation, credit is given for all such responses, regardless of whether or not they are paired with the appropriate stimulus; the same measure can be obtained for second-list responses. A parallel procedure can be used to assess the availability in free recall of items from two or more successive lists. In a study by Tulving and Thornton (1959), which was concerned with the interaction of retroactive and proactive effects, the number of successive word lists was varied from one to five, and recall of all the lists learned by a given group was required on the terminal test. The results showed progressive decreases in first-list recall as a function of the number of interpolated lists. Using nonsense syllables, Asch and Ebenholtz (1962) found significant RI when *S*s were instructed to recall both the original and the interpolated list. With nonsense materials, however, there is an inevitable overlap of elements in the successive lists, and it is not possible to assess the relative amounts of retention loss attributable to decrements in response integration on the one hand, and to reduced availability of intact units on the other. The present study of RI in free recall focuses on the latter question and uses lists of words as the materials. Degree of IL was manipulated since this variable is known to be a major determinant of the amount of unlearning in the paired-associate situation (Barnes & Underwood, 1959).

METHOD

Design.—The design comprised three experimental groups and a control group. The experimental groups learned two 20-word lists in succession. There were four trials of free-recall learning on the first list, and either two, four, or six trials on the second list. After the end of second-list learning retention of both lists was tested. The control group had four learning trials and a recall test on a single list. The experimental treatments will be designated as Cond. IL_2, IL_4, and IL_6, respectively, and the control treatment as Cond. IL_0.

Materials.—There were two lists of 20 two-syllable nouns each. The frequencies in the L count of Thorndike and Lorge (1944) were in the range from 100 to 333 for half the words in each list, and in the range from 10 to 33 for the other half. There were either five or six duplications of first letters per list. All but one of the first letters were used in both lists. Meaningful similarity was minimized both within and between lists. Each list was used equally often as the original and as the interpolated task.

Procedure.—The learning procedure consisted of alternate study and test trials. On the study trials the words were read aloud by *E* at the rate of one per sec. Immediately after the last word a bell was rung to signal the beginning of a 90-sec. test trial. The *S*s were instructed to write down all the words they could remember in any order they wished. For purposes of recording their responses *S*s were provided with a set of consecutively numbered cards one of which was used for each test. A bell was rung twice to indicate the end of the test period, at which time *S* placed the completed card at the bottom of his pile. There was a 10-sec. interval between each test trial and the next study trial. Under the experimental conditions the interval between the last test trial on the first list and the first study trial on the second list was 40 sec.

A different random order of presentation was used on each study trial, subject to the restriction that no word occupied the same serial position more than once.

Under the experimental conditions a terminal 5-min. test of free recall for both lists was given 40 sec. after the end of the last test trial on the second list. Recall instructions were given during the 40-sec. interval. The *S*s were asked to write all the words they could remember in any order they wished, and it was stated that the items need not be grouped by list. At the end of the recall period *S*s were instructed to identify the list membership of all the words they had written down by writing the no. 1 or 2 next to each item. Under the control condition (IL_0) *S*s worked on a series of mathematical problems for a period of time equal to that occupied by second-list learning under Cond. IL_6; the retention interval was 13 min. The instructions were, of course, modified so as to refer to only one list. The procedure of administering the test of recall immediately after the end of IL follows that of Barnes and Underwood (1959) and has the advantage of eliminating the possibility of spontaneous recovery. Consequently the interval between the end of first-list learning and the test of recall increases with the number of trials of IL. The use of the longest of these intervals under the control condition provides a conservative baseline for the evaluation of the amount of RI.

Subjects.—Each condition was administered to six groups of *S*s varying in size between five and nine. The assignment to conditions was made in six blocks, with the four conditions represented once in each block. The running order within blocks was determined by means of a table of random numbers. In order to obtain equal *N*s, each of the six groups within a condition was reduced to five by random elimination of excess protocols prior to scoring. The analysis of the results, is, therefore, based on 30 *S*s (six groups of five *S*s each) per condition. All *S*s were undergraduate students at the University of California who were not necessarily naive to rote-learning experiments.

RESULTS

First-list learning.—The means of the total numbers of correct responses summed over the four test trials of first-list learning were as follows: IL_0—45.8, IL_2—47.2, IL_4—45.4, IL_6—46.4. The amounts of first-list learning under the four conditions are, therefore, closely comparable, $F < 1$. The mean number correct on the last test trial of first-list learning were 14.3, 14.6, 14.3, and 14.2, respectively, for the four conditions considered in the same order as above.

Second-list learning.—The mean number of correct responses in List-2 learning were: IL_2—19.0, IL_4—46.6, IL_6—78.1. The mean scores on the final test trial of IL were 11.5, 14.6, and 15.4, respectively. For the final scores, $F\ (2, 87) = 18.59$, $p < .001$. Further analysis shows that Cond. IL_4 and IL_6 reach a significantly higher level of performance than Cond. IL_2 ($p < .001$) but do not differ reliably from each other. It is apparent that IL proceeds at a negatively accelerated rate.

Intrusions of List-1 responses during List-2 learning were infrequent but showed some increases as a function of successive trials within conditions. The total frequencies of such errors, with the numbers of *S*s contributing them given in parentheses, were as follows: IL_2—1 (1), IL_4—12 (6), IL_6—15 (5).

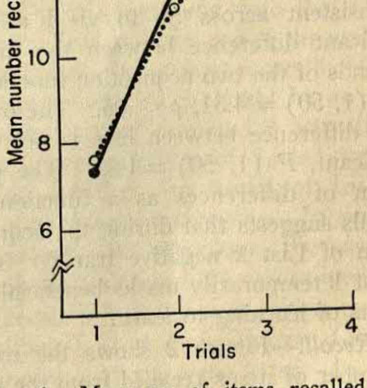


FIG. 1. Mean no. of items recalled on successive trials of List-1 and List-2 learning (Cond. IL_4 and IL_6).

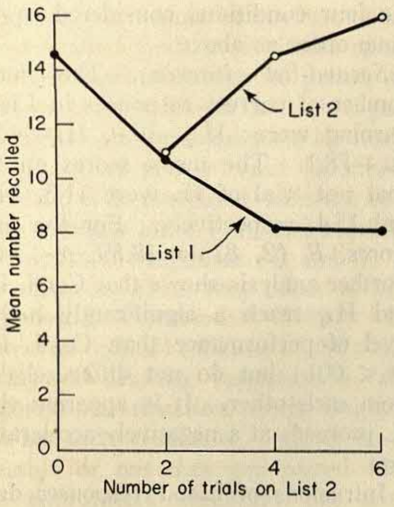


FIG. 2. Mean no. of items recalled from the two lists on the terminal test of retention.

Transfer effects.—Conditions IL_4 and IL_6 permit an assessment of net transfer effects in IL as measured by the differences between the levels of performance on corresponding trials on List 1 and List 2. The scores for Trials 1–4 on the two lists are shown in Fig. 1. While the separations between the two curves are small, a systematic trend is nevertheless indicated. On Trial 1, List 2 falls slightly below List 1; thereafter List 2 has an advantage. This pattern is sufficiently consistent across *S*s to yield a significant difference between the linear trends of the two acquisition functions, F (1, 50) = 4.31, $p < .05$. The overall difference between lists is not significant, F (1, 50) = 1.85. The pattern of differences as a function of trials suggests that during the acquisition of List 2 negative transfer from List 1 temporarily masks beneficial effects of learning to learn.

Recall.—Figure 2 shows the mean number of items recalled from the two lists on the terminal test as a function of the number of trials on List 2. Recall of the first list decreases sharply after two trials of IL and more gradually thereafter. Recall of the second list shows progressive increases. The changes in the recall of the original and the interpolated materials exhibit a pattern quite similar to that observed with paired associates (Barnes & Underwood, 1959). The differences among conditions in List-1 recall are highly significant, F (3, 116) = 22.96, $p < .001$. Orthogonal comparisons show that the difference between the control condition and the combined experimental conditions is significant at the .001 level, and the variation among the latter at the .01 level. Thus, there is significant RI which varies reliably with the number of trials of IL.

Identification of the list membership of the correct responses was uniformly high. The percentages of correct identifications of List-1 items were 92.0, 92.6, and 94.0 for Cond. IL_2, IL_4, and IL_6, respectively. The corresponding percentages for List-2 items were 95.2, 94.4, and 98.0.

Importations of responses from outside the list were rare. The total frequencies of such errors, with the numbers of *S*s contributing them in parentheses, were as follows: IL_2—16 (10), IL_4—11 (9), IL_6—10 (8).

Order of recall.—In MMFR tests for paired associates, there is an increasing tendency as a function of the number of IL trials for List-2 response to precede List-1 responses in recall (Barnes & Underwood, 1959). To determine whether there was a similar trend in the present data the following procedure was used. The percentage of List-2 items in the first half of the words recalled was first determined for each *S*. If the order of recall were entirely random, this percentage should be equal to the percentage of List-2 items in all the correct responses. The difference between the two percentages thus defines an index of List-2 priority in recall: the larger the difference the greater the priority. To take account of variations in expected values, the obtained difference was expressed as a percentage of the possible difference. The mean values of the adjusted index are 23.8, 30.7, and 50.6 and clearly show a progressive increase in List-2 dominance as a function of the number of trials of IL. Analysis of the adjusted measures (after arc-sine transformation) yields a significant linear trend, $F\ (1,\ 87) = 4.18$, $p < .05$.

Output interference.—Given the fact that List-2 items were likely to be recalled before List-1 items, the question arises whether and to what extent output interference contributes to the observed RI. That is, is there interference with List-1 recall by intervening reproductions of List-2 items (cf. Tulving & Arbuckle, 1963)? This question cannot be answered directly from the present data. However, on the assumption that output interference had a significant effect, there should be an inverse relationship between the amounts recalled from List 1 and from List 2 when degree of learning for the two lists is held constant. To test this implication, second-order partial correlations between List-1 and List-2 recall were computed; the measures of degree of learning which were held constant were the total numbers of correct responses during the acquisition of each list. The coefficients of partial correlation are $-.39$ for IL_2 ($p < .05$), $+.53$ for IL_4 ($p < .01$), and $+.27$ for IL_6 ($p > .05$). The finding of interest is the change from a negative relationship for IL_2 to a positive one for IL_4 and IL_6, indicating a shift from net output interference to facilitation.

One possible interpretation of the pattern of correlations takes its point of departure from the phenomenon of subjective organization, i.e., the development of higher-order units in free-recall, described by Tulving (1962b; in press). It may be assumed that as IL progresses at least some of the items in List 2 are grouped into categories similar to those used during the acquisition of List 1. Once such categorical groupings have become stabilized, i.e., at the higher levels of IL, the reproduction of items from List 2 serves to facilitate the recall of related items from List 1 which have remained available, and conversely. On this assumption there is net output interference only before stable groupings of the List-2 items have been developed, i.e., at the lowest level of IL. The relatively low positive correlation for Cond. IL_6 may be due to the fact that recall of List 2 approaches its ceiling.

A source of organization which is likely to be common to the successive lists is alphabetical grouping. Such grouping can be achieved readily and is known to facilitate recall (Tulving, 1962a). According to the present analysis of interlist facilitation in recall, there should be an increasing correlation as a function of the degree of IL between the recalls of List-1 and List-2 words beginning with the same letter. That is, if the use of the alphabetical category becomes more likely as IL progresses, there should be an increasing probability that the activation of a given category of List-2 items should facilitate the recall of List-1 items in the same category. In order to evaluate this prediction, the conditional probabilities were determined of the recall of List-1 words, given recall of either 0, 1, or 2 words from List 2 beginning with the same letter. In the calculation of these probabilities, the frequencies of recall were pooled over

TABLE 1

CONDITIONAL PROBABILITIES OF RECALL OF LIST–1 ITEMS, GIVEN DIFFERENT FREQUENCIES OF RECALL OF LIST–2 ITEMS BEGINNING WITH THE SAME LETTER

No. List-2 Items	Cond.		
	IL_2	IL_4	IL_6
0	.495	.328	.264
1	.549	.428	.394
2	.549	.429	.449

*S*s. (As was stated earlier, there was only one first letter which was not common to the two lists and which had therefore to be omitted from the present analysis.) The conditional probabilities are presented in Table 1. The relationship between List-1 and List-2 recall of words beginning with the same letter is slight in Cond. IL_2 but increases with the number of trials of IL; it is clearly present in Cond. IL_6. It should be noted that the level of conditional probabilities declines with the degree of IL whereas the apparent interdependence of List-1 and List-2 recall increases. Thus, interlist facilitation appears to increase even though the absolute availability of List-1 responses declines.

Recall as a function of degree of original learning.—A substantial proportion of items remains resistant to interference even when List 2 is practiced for more trials than List 1 (Fig. 2). An item analysis was carried out to determine whether differential resistance to interference is related directly to ease of acquisition. For purposes of this analysis, the 10/20 items recalled most frequently during the test trials of List-1 learning were designated as easy, and the remaining 10/20 items as difficult. The percentage of easy items to the base of all List-1 items recalled on the terminal test was then determined for each *S*; the higher the percentage the greater is the predominance of easy items in recall. The mean percentages were as follows: IL_0—60.4, IL_2—65.2, IL_4—65.9, IL_6—71.2. The relative predominance of easy items increases with the degree of IL. The linear trend is significant, $F(1, 116) = 10.51$, $p < .01$ after arcsine transformation. For each of the conditions, the mean percentage exceeds 50 at the .001 level.

Sequential order in recall.—The resistance of List-1 items to interference may be enhanced by membership in stable higher-order units. On this assumption clusters developed during acquisition should be preserved in recall; moreover, the heavier the interference the greater should be the preponderance of stable clusters among the items recalled. Pairs of items occurring consecutively both on the last trial of acquisition and on the final test of recall probably represent instances of such clusters. The number of such repeated pairs was, therefore, determined for each *S*. In the analysis of these results the method developed by Bousfield, Puff, and Cowan (1964) was used which calls for an evaluation of the difference between the number of obtained intertrial repetitions (O-ITRs) and the number of such repetitions expected by chance (E-ITRs). The relevant measures for the recall of List-1 items are presented in Table 2. For each of the four conditions the obtained values exceed the expected values, and in each case the difference is significant at the .01 level. In order to make possible a comparison of the conditions which is independent of the absolute levels of recall (n), the numbers of O-ITRs were expressed as percentages of the maximal possible values ($n - 1$). In the calculation of these scores both the obtained and the maximal values were corrected for chance.

TABLE 2

Mean Measures of Intertrial Repetition of Pairs: Last Trial of Acquisition vs. Test of Recall

	List 1				List 2		
	IL_0	IL_2	IL_4	IL_6	IL_2	IL_4	IL_6
E-ITR	.49	.36	.27	.27	.29	.48	.59
O-ITR	2.17	1.83	.97	1.07	1.37	2.17	3.10
O-E	1.68	1.47	.70	.80	1.08	1.69	2.51
Percentage	12.44	15.00	8.95	9.52	10.40	12.76	17.52

Note.—E-ITR = expected intertrial repetition; O-ITR = obtained intertrial repetition; percentage is to the base of maximum possible repetition score corrected for chance.

The mean percentages, which thus represent relative measures of ITR, are shown in Table 2. There is a slight increase between IL_0 and IL_2, followed by a decline. However, the variation among conditions is not significant, $F = 1.04$.

The results of a parallel analysis for List-2 items are also presented in Table 2. Again the difference between the obtained and expected value is significant at the .01 level in each case. Note the virtual identity of the measures for List-1 recall in Cond. IL_0 and List-2 recall in Cond. IL_4. In both cases there were four trials of acquisition followed by a test of retention without an interpolated task. The percentage measures for List 2 increase with the number of trials of IL, suggesting progressive increases in the stability of the units, but these differences also fail to reach significance, F (2, 87) = 1.96.

Discussion

When a test analogous to MMFR is used to measure RI in free recall, the manifestations of interference parallel those observed for paired associates. There is a negatively accelerated increase in RI as a function of degree of IL, accompanied by a progressive tendency for List-2 items to precede List-1 items in recall; identification of list membership, however, remains highly accurate throughout. Retention losses in MMFR have been used to define the amount of unlearning for paired associates. To the extent that the test used in the present study and MMFR are, indeed, functionally equivalent, the results add to the evidence for unlearning in the free-recall situation (cf. Asch & Ebenholtz, 1962; Tulving & Thornton, 1959). A first step in the application of the theoretical concepts developed in the analysis of paired-associate studies would be to attribute unlearning in free recall to the loss of contextual associations. Thus, the losses found in the A-B, C-D paradigm and in free recall would be subsumed under a common principle. Such an assumption would be supported first of all by considerations of parsimony. In addition, there are a priori arguments for the dependence of free recall on the integrity of contextual associations. Such associations or their functional equivalents appear to be required to account for the initiation of specific recall sequences; hence the disruption of contextual associations is likely to contribute to losses in free recall. However, to equate unlearning in free recall with the loss of contextual associations would probably represent an incomplete interpretation.

There is considerable evidence for a positive relationship between strength of interitem associations and indexes of subjective organization on the one hand and amount of free recall on the other (e.g., Deese, 1959, 1960; Tulving, 1962b). Hence the unlearning of interitem associations and/or the disruption of higher-order units must be considered as a potential condition of RI in free recall. It must be acknowledged that the analysis of ITRs in the present experiment does not lend clear support for the operation of such a process. While there is statistically reliable evidence for the persistence of sequential groupings between the end of acquisition and the final test of recall, the absolute numbers of ITRs are low and their relative frequency of occurrence does not vary significantly as a function of the degree of IL. These results must, however, be interpreted with caution for two reasons. (*a*) The validity of ITRs as measures of stable sequential groupings is uncertain. (*b*) The persistence of *sequential order* is only one possible manifestation of the influence of interitem linkages. For example, *S*s using categorical groupings as aids to recall may review a series of categories repeatedly during a single test trial; under these circumstances the effectiveness of the categories could not be assessed from the order of recall. Furthermore, when there are multiple associative connections among the members of a series, the activation of alternative sequences will also make for variability in the order of recall.

The point that consistency of the order of recall is only one possible index of associative relations among items is worthy of emphasis in view of the critical significance which Asch and Ebenholtz (1962) assign to their failure to find clear evidence for such consistency in the free recall of lists of nonsense syllables. In the light of other evidence mentioned above for the effectiveness of interitem associations the finding of variability in the order of recall does not by itself support the conclusion of these investigators that free recall is a nonassociative process.

It was suggested earlier that conditions of negative transfer may obtain for specific interitem associations during the acquisition of successive lists by free learning. The evidence for interlist facilitation in recall at the higher levels of IL indicates that for some items, at least, the functional conditions of transfer may be different during acquisition and on the test of retention. (For a discussion of the distinction between nominal and functional conditions of transfer see Postman, 1963, pp. 44–45.) When items from the two lists are categorized in the same way, the A-B, A-C paradigm will apply during List-2 learning, e.g., new responses may be attached to common category names. On the final test, however, the recall of List-1 items which have resisted unlearning will be facilitated by the reproduction of related words from List 2. For the items available at the time of the test the conditions of performance may be represented in terms of mediated sequences like A-B–A-B′, etc. It may be noted that such groupings of items from the two lists will also detract from the consistency of the order of recall of first-list items.

In conclusion, successively higher degrees of IL appear to produce similar changes in free recall and in paired-associate recall. The presence of functional similarities does not, of course, necessarily imply an identity of mechanisms. However, the results lend some encouragement to the view that unlearning is a general principle of interference which is not contingent on a particular method of practice. This view is also supported by the evidence for unlearning in the recall of serial lists (Keppel, 1966). An interpretation of the present results in terms of unlearning of both contextual and sequential associations also appears to be consistent with the available evidence on the conditions and characteristics of free-recall learning. This interpretation differs from that of Asch and Ebenholtz (1962) who attribute RI in free recall to a reduction in item availability as a result of interpolation. As Tulving (in press) has pointed out, such a formulation fails to specify the mechanisms governing the reproduction of available items, i.e., the initiation and maintenance of recall sequences. In the analysis of these mechanisms the role of associative processes deserves continued attention.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. The process of free recall: Evidence for non-associative factors in acquisition and retention. *J. Psychol.*, 1962, **54**, 3–31.

Barnes, J. M., & Underwood, B. J. 'Fate' of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during free recall. *J. verbal Learn. verbal Behav.*, 1964, **3**, 449–459.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

Deese, J. Frequency of usage and number of words in free recall. The role of association. *Psychol. Rep.*, 1960, **7**, 337–344.

Keppel, G. Unlearning in serial learning. *J. exp. Psychol.*, 1966, **71**, 143–149.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychol. Monogr.*, 1964, **78**(16, Whole No. 593).

Postman, L. Does interference theory predict too much forgetting? *J. verbal Learn. verbal Behav.*, 1963, **2**, 40–48.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Tulving, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canad. J. Psychol.*, 1962, **16**, 185–191. (a)

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–354. (*b*)

Tulving, E. Free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior theory and its relation to general S-R theory.* New York: Prentice Hall, in press.

Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. *J. verbal Learn. verbal Behav.*, 1963, **1**, 321–334.

Tulving, E., & Thornton, G. B. Interaction between proaction and retroaction in short-term retention. *Canad. J. Psychol.*, 1959, **13**, 255–265.

(Received June 22, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 212–218

LEARNING SEVERAL CONCEPT-IDENTIFICATION PROBLEMS CONCURRENTLY: A TEST OF THE SAMPLING-WITH-REPLACEMENT ASSUMPTION[1]

JAMES R. ERICKSON AND MYRON M. ZAJKOWSKI

Ohio State University

48 Ss learned 3 concept-identification problems concurrently. 1 was a control problem, 1 involved reversal shifts, and the 3rd involved non-reversal shifts after every other error. Choice data conformed to predictions from all-or-none models which assume sampling with replacement. Latency data showed that latencies after errors (AE) were longer than after corrects (AC), AC latencies declined during the presolution period but AE latencies did not, and postcriterion latencies showed a "learning curve." Results were compared with those where the same problems were learned successively. It was suggested that the concurrent procedure leads to interference with S's short-term memory of hypotheses tried but rejected which causes data to conform more closely with predictions from sampling-with-replacement models.

Recent studies have called into question one of the basic assumptions of all-or-none theories of concept identification (Bower & Trabasso, 1964; Restle, 1962). The assumption which seems to prove the most troublesome is that of "sampling with replacement," i.e., the assumption that, following an error, *S*s sample with replacement from a pool of hypotheses, having no memory of hypotheses tried but rejected. Among data questioning this assumption are effects of random reinforcement (Holstein & Premack, 1965; Levine, 1962), questioning *S*s about hypotheses used (Trabasso & Bower, 1964a), analysis of additivity of cues (Trabasso & Bower, 1966), and analysis of latency data (Erickson, Zajkowski, & Ehmann, 1966).

To account for these data, Erickson et al. (1966) and Trabasso and Bower (1966) have proposed that *S*s do not sample with replacement, but retain some information (either specific stimulus response pairings or hypotheses tried but rejected) from past trials for a period of time in short-term memory until it is replaced by more recent information. The present experiment was designed to give further information on *S*'s memory while solving concept-identification problems. In this experiment the information load was increased by having *S*s work concurrently on three concept-identification problems. This procedure should overload short-term memory making *S*'s performance more in line with the assumptions of sampling-with-replacement models, since as the information load becomes larger it should be more difficult to retain either specific stimulus-response assignments or past hypotheses tried from any one of the problems because of interference from interpolated instances of the other two problems. This procedure has recently been used successfully by Restle and Emmerich (1966). It was therefore hypothesized that the choice and latency data would follow predictions

---

[1] This research was carried out at the Human Performance Center and was supported in part by National Institutes of Health Grant No. MH-12113-01. A paper based on this investigation was read at the 1966 Midwestern Psychological Association meetings in Chicago.

TABLE 1

Summary Error Data for Three Reversal Conditions

Cond.	n	Mean No. Errors (T)	*SD*	Predicted Mean TLE	Obtained Mean TLE	*SD*
Control	48	5.71	6.99	11.42	12.10	15.14
Reversal Shifts	48	5.33	4.41	19.55	19.33	17.33
Nonreversal Shifts	48	4.62	4.33	16.48	17.42	17.26

from all-or-none, sampling-with-replacement models.

METHOD

Subjects.—The *S*s were 57 students who were recruited from the introductory psychology *S* pool. Participation in experiments was a course requirement. Of these *S*s, nine were not able to solve one or more of the three problems within the experimental session and were excluded from the analysis. The design called for 48 *S*s divided evenly among three conditions, so, as *S*s were excluded, they were replaced until all conditions were filled.[2]

Procedure.—Each *S* received fairly detailed instructions concerning the nature of the task—emphasizing that the solution to each problem involved only one dimension, that *E* would assign Responses A and B to the two values of the correct dimension, and that *S*'s task was to discover *E*'s rule for response assignment. Two practice problems, each involving two binary dimensions, were then solved by *S*. Before solving the main problems, the several dimensions were pointed out, and *S*s were allowed to examine sample stimulus materials.

Each *S* solved three main problems: One was a control, where the assignment of responses was constant through the learning period; one involved reversals of response assignments within the same dimension after every other error; and the third involved nonreversal shifts to another dimension after every other error—the procedure was as described by Bower and Trabasso (1963).

The three concept-identification problems each contained four binary dimensions. The problems were called the square problem, the flower problem, and the letter problem, after their most salient features—examples of these stimuli may be found in Erickson et al. (1966). The various stimuli were placed on 35-mm. slides and rear projected onto a screen in front of *S*. The *S*'s booth contained, besides the screen, a response box with two buttons and two feedback lights, labeled "A" and "B." Response latency was measured from the time of slide-onset to button-press by a 1/100-sec. Standard Electric timer.

Feedback followed the button press by 2 sec. and remained on for 1 sec., after which both the feedback light and the slide went off. Ten seconds later the next slide came on. The *S*s wore earphones into which low-level white noise was projected to eliminate distracting sounds from the projection equipment, timers, relays, etc.

The assignment of reversal conditions to stimulus materials was counterbalanced across *S*s, so for example, 16 *S*s solved the flower problem as a control, the square problem as a reversal-shift problem, and the letter problem as the nonreversal-shift problem. The correct solution for each problem (and the dimension shifted to for the nonreversal problem) was randomly selected for each *S*, with the restriction that two *S*s received each of the eight possible solutions as a starting point for each problem within each 16-*S* condition.

Stimulus materials were placed in several random orders, with the restriction that every 48 slides contained all of the possible stimuli (16 from each problem). The learning criterion was 20 successive correct responses on each problem—problems were *not* dropped out as they were solved, so *S*s continued to respond to all three problems until the last one reached criterion (*S*s were informed of the criterion). All three problems were solved in a single experimental session which usually lasted about 1½ hr.

RESULTS

Frequency data.—The mean number of informed errors (*T*) for the three reversal conditions is given in Table 1. Analysis of variance of these data revealed no significant differences among the conditions, $F\ (2, 94) = .61$, $p > .50$. These data provide another replication of the Bower and Trabasso (1963) results Trabasso and Bower, 1964b in a situation where each *S* not only acts as his own control, but solves all three types of problems concurrently. It might be noted that these data, like most, show a slight nonsignificant (by orthogonal contrast) tendency for the nonreversal problem to be solved with fewer errors than the other problems. This tendency has been predicted by Trabasso and Bower (1966) in a recent article.

---

[2] Of the 27 problems attempted by the nine rejected *S*s, two control problems, eight reversal problems, and five nonreversal problems were not solved.

The trial of last error (TLE) can be predicted from all-or-none models from the total error data, with the assumption that the probability of an er-

TABLE 2

COMPARISON OF OBSERVED DATA FOR CONTROL PROBLEMS WITH PREDICTIONS FROM THE BOWER AND TRABASSO MODEL

<table>
<tr><th>Problem</th><th colspan="6">Square</th><th colspan="6">Flower</th><th colspan="6">Letter</th><th colspan="6">Combined</th></tr>
<tr><td>ϵ</td><td colspan="6">.235</td><td colspan="6">.139</td><td colspan="6">.176</td><td colspan="6">.175</td></tr>
<tr><td>p</td><td colspan="6">.614</td><td colspan="6">.496</td><td colspan="6">.474</td><td colspan="6">.527</td></tr>
<tr><td>n</td><td colspan="6">16</td><td colspan="6">16</td><td colspan="6">16</td><td colspan="6">48</td></tr>
<tr><td></td><td colspan="24">Precriterion Tests[b]</td></tr>
<tr><td></td><td colspan="2">df</td><td colspan="2">χ^2</td><td colspan="2">p</td><td colspan="2">df</td><td colspan="2">χ^2</td><td colspan="2">p</td><td colspan="2">df</td><td colspan="2">χ^2</td><td colspan="2">p</td><td colspan="2">df</td><td colspan="2">χ^2</td><td colspan="2">p</td></tr>
<tr><td>Stationarity</td><td colspan="2">4</td><td colspan="2">2.67</td><td colspan="2">>.60</td><td colspan="2">5</td><td colspan="2">3.87</td><td colspan="2">>.50</td><td colspan="2">4</td><td colspan="2">2.09</td><td colspan="2">>.70</td><td colspan="2">7</td><td colspan="2">6.82</td><td colspan="2">>.40</td></tr>
<tr><td>Independence</td><td colspan="2">1</td><td colspan="2">.977</td><td colspan="2">>.30</td><td colspan="2">1</td><td colspan="2">2.809</td><td colspan="2">>.05</td><td colspan="2">1</td><td colspan="2">.187</td><td colspan="2">>.60</td><td colspan="2">1</td><td colspan="2">4.873</td><td colspan="2"><.05</td></tr>
<tr><td>Binomial</td><td colspan="2">4</td><td colspan="2">3.17</td><td colspan="2">>.50</td><td colspan="2">4</td><td colspan="2">2.27</td><td colspan="2">>.60</td><td colspan="2">4</td><td colspan="2">1.14</td><td colspan="2">>.80</td><td colspan="2">4</td><td colspan="2">6.08</td><td colspan="2">>.30</td></tr>
<tr><td></td><td colspan="24">Other Statistics</td></tr>
<tr><td></td><td colspan="3">Obs.</td><td colspan="3">Pred.</td><td colspan="3">Obs.</td><td colspan="3">Pred.</td><td colspan="3">Obs.</td><td colspan="3">Pred.</td><td colspan="3">Obs.</td><td colspan="3">Pred.</td></tr>
<tr><td>Mean No. Errors (T)[a]</td><td colspan="3">4.250</td><td colspan="3">—</td><td colspan="3">7.188</td><td colspan="3">—</td><td colspan="3">5.688</td><td colspan="3">—</td><td colspan="3">5.708</td><td colspan="3">—</td></tr>
<tr><td>*SD*</td><td colspan="3">2.769</td><td colspan="3">3.716</td><td colspan="3">9.799</td><td colspan="3">6.669</td><td colspan="3">6.680</td><td colspan="3">5.163</td><td colspan="3">6.986</td><td colspan="3">5.184</td></tr>
<tr><td>Trial of Last Error (TLE)</td><td colspan="3">11.000</td><td colspan="3">11.000</td><td colspan="3">14.500</td><td colspan="3">14.252</td><td colspan="3">10.812</td><td colspan="3">10.812</td><td colspan="3">12.104</td><td colspan="3">12.060</td></tr>
<tr><td>*SD*</td><td colspan="3">6.947</td><td colspan="3">10.488</td><td colspan="3">22.815</td><td colspan="3">13.743</td><td colspan="3">11.839</td><td colspan="3">10.300</td><td colspan="3">15.139</td><td colspan="3">11.549</td></tr>
<tr><td>No. Correct Prior to Last Error</td><td colspan="3">6.750</td><td colspan="3">6.750</td><td colspan="3">7.188</td><td colspan="3">7.065</td><td colspan="3">5.125</td><td colspan="3">5.125</td><td colspan="3">6.375</td><td colspan="3">6.352</td></tr>
<tr><td>*SD*</td><td colspan="3">5.000</td><td colspan="3">7.233</td><td colspan="3">12.993</td><td colspan="3">7.548</td><td colspan="3">5.932</td><td colspan="3">5.603</td><td colspan="3">8.712</td><td colspan="3">6.834</td></tr>
<tr><td>No. Correct between *k*th and *k* + 1st Error</td><td colspan="3">1.588</td><td colspan="3">1.588</td><td colspan="3">1.017</td><td colspan="3">.983</td><td colspan="3">.901</td><td colspan="3">.901</td><td colspan="3">1.120</td><td colspan="3">1.113</td></tr>
<tr><td>*SD*</td><td colspan="3">2.211</td><td colspan="3">2.027</td><td colspan="3">1.560</td><td colspan="3">1.396</td><td colspan="3">1.407</td><td colspan="3">1.309</td><td colspan="3">1.722</td><td colspan="3">1.533</td></tr>
<tr><td>Error-Correct Alternations</td><td colspan="3">5.188</td><td colspan="3">5.602</td><td colspan="3">6.938</td><td colspan="3">7.630</td><td colspan="3">5.750</td><td colspan="3">5.918</td><td colspan="3">5.958</td><td colspan="3">6.486</td></tr>
<tr><td>No. Error Runs</td><td colspan="3">2.312</td><td colspan="3">2.994</td><td colspan="3">3.188</td><td colspan="3">4.067</td><td colspan="3">2.625</td><td colspan="3">3.222</td><td colspan="3">2.708</td><td colspan="3">3.480</td></tr>
</table>

[a] The estimate of ϵ is $1/T$, so there can be no predicted value for this statistic.
[b] These tests are explained in detail in Suppes and Ginsberg (1963).

ror during the presolution period remains constant at .50. The formulas are TLE = 2T for the control problem and TLE = 4T − 2 for the reversal problems, where T is the number of informed errors for the various problems. Obtained values were not significantly different from predicted values, t (47) = .31, $p > .75$ for the control problem, t (47) = .09, $p > .90$ for the reversal problem, and t (47) = .38, $p > .75$ for the nonreversal problem. Means and *SD*s for TLE are given in Table 1.

Other aspects of the frequency data for the control problems are summarized in Table 2. The predictions included in the table are from the Bower and Trabasso (1964) theory.

It can be seen from Table 2 that there was little, if any, evidence of learning before TLE. During these trials the probability of a correct response did not increase significantly across trials (stationarity); for the first Vincentized half of the precriterion trials, p (C) = .525 and for the second half p (C) = .532. The probability of an error was just as large after an error as after a correct response (independence) although when the three problems were combined, $p(E_n|E_{n-1}) = .52$ and $p(E_n|C_{n-1}) = .42$; these were significantly different, $p < .05$. When trials were grouped in blocks of four, the distributions of correct responses closely approximated binominal distributions with p values given by the proportion of correct responses over the precriterion period and, when the precriterion trials were Vincentized; the blocks of four distributions for the two halves did not differ significantly, $\chi^2(3) = 3.18$, $p > .30$.

Point predictions of several statistics involving parameters estimated from the data show excellent fits for mean predictions, and fair fits for predictions of variability. The model has not been worked out in detail for problems involving shifts, so these problems cannot be examined in as much detail.

Latency data.—The latency data were first examined as a function of problem type and trial type. Latencies on trials following errors (AE), following correct responses (AC), on error trials (OE), and on correct trials (OC) were all examined during the presolution period. (For example, when an error was made on the control problem, the AE latency was noted on the next trial of that problem, ignoring intervening trials on the other problems.)

For the AE vs. AC data, median latencies for each *S* problem were calculated for all *S*s who had at least one error and one correct precriterion response on each problem (a total of 23). These medians were used as the dependent variable for the analysis of variance. Latencies after errors were significantly longer than latencies after correct responses, F (1, 22) = 16.10, $p < .001$. Latencies did not differ among the three problem types, F (2, 44) = 2.13, $p > .10$, nor was there any interaction between trial type and problem type, F (2, 44) = .31, $p > .70$. A similar analysis of OE and OC latencies revealed no main effects which were significant, but the interaction between trial type and problem type reached significance, F (2, 56) = 4.03, $p < .05$. This interaction occurred because OE latencies were slightly longer than OC latencies for the control and reversal problems, while the reverse was true for the nonreversal problem. Since similar interactions have not occurred in other studies, perhaps the most parsimonious interpretation is to assume that this interaction represents a Type I error—at least until it is replicated.

A graphic presentation of latencies before and after TLE is presented in Fig. 1. The values of the precriterion latencies must be interpreted carefully since at any trial number, those *S*s represented could have just made their first, second, etc., error—the only thing in common is that they made an error on the previous trial. Nevertheless, the difference between AE and AC latencies is clear. Also clear is a "learning curve" for latencies following the last error—this has been found previously (Erickson et al., 1966; Millward, 1964). One interpretation of this finding was given by Erickson et al. (1966) who suggest that this decline is caused by *S*s focusing on one aspect of the stimulus as he gains confidence that his hypothesis is the correct one. Latencies in the presolution period also seem to decline—latencies closer to the TLE are shorter than latencies earlier in learning.

An analysis of variance was performed to confirm this—for each *S*, AE and AC trials were split into two halves, and median latencies were computed for each half. The mean of these median latencies across problems was the dependent variable in the analysis; thus, each *S* who had at least two AE and two AC trials on at least one of his problems could be included in the analysis—43 of the 48 *S*s met this criterion. The analysis showed that the main effect of trial type and the interaction between halves and trial type were significant, $F\ (1, 42) = 26.82,\ p < .001$ and $F\ (1, 42) = 7.99,\ p < .01$, respectively, while the main effect of halves was not significant, $F\ (1, 42) = .93,\ p > .25$. Means of the various conditions are given in Table 3. It seems that the decline in AE latency in Fig. 1 might be artifactual, being primarily a between-*S* effect, since in the median data, there was a slight increase in latency from the first to the second half of the precriterion trials. For AC trials, however, latency declined over ½ sec. from the first to the second half.

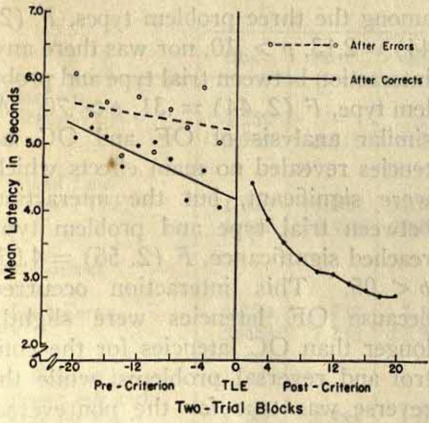


FIG. 1. Mean latency on trials preceding and following TLE. (Precriterion trials show latencies on Trial *n* following errors and following correct responses on Trial $n - 1$ for all problems and all *S*s combined. Best-fit linear regression lines are shown for reference.)

TABLE 3

MEAN OF MEDIAN LATENCIES (IN SEC.) AFTER ERRORS AND AFTER CORRECT RESPONSES FOR THE FIRST AND SECOND HALVES OF THE PRECRITERION PERIOD

	AE	AC	*M*
1st Half	4.734	4.501	4.618
2nd Half	4.957	3.980	4.468
M	4.845	4.241	4.543

DISCUSSION

The implications of these data become more clear when they are compared with corresponding data for the same problems learned under similar conditions—except learned successively instead of concurrently by *S*. The successive data are presented in Erickson et al. (1966).

It is clear that learning these problems simultaneously makes them more difficult—an average of 1.6 more errors were made per problem (5.23 vs. 3.63).[3] This difference is, of course, reflected in estimates of ϵ, the learning rate parameter for the control problems—.175 for the simultaneous data vs. .292 for the successive data. If *S*s were using each possible response assignment as a hypothesis and sampling with replacement from the pool, the a priori value of ϵ would be $\frac{1}{8}$–.125; it is clear that the simultaneous procedure moves ϵ towards this value.

If *S*s were treating the problems as having eight independent hypotheses (rather than four hypothesis with two possible S-R connections per hypothesis, as the Bower and Trabasso model suggests), an equal number of *S*s in the reversal and nonreversal conditions should be "switched in" to a solution, i.e., have an uninformed error following their last informed error. For the simultaneous condition 13 reversal problems were solved in this manner vs. 16 nonreversal problems (out of 48). In the successive data, for 9 reversal problems vs. 18 nonreversal problems (out of 54), the last error was uninformed. Trabasso and Bower (1964b) state that for their *S*s (who solved one problem only with all *S*s solving on the same dimensions) the corresponding numbers were 1 out of 26 for reversal vs. 13 out of 26 for nonreversal.

The latency data also reflect the increased difficulty of the concurrent method. The mean latency during the presolution period under these conditions was .4–.8 sec. longer (depending on the method of analysis, by trials or by halves), and asymptotic postcriterion latencies were .63 sec. longer than under the successive method. This mean increase undoubtedly reflects two things, increased time to recognize or classify the stimulus type and increased time to try to recall the solution for that particular stimulus type because of the memory load.

The second factor becomes clear if the AE and AC data are examined across trials. With the successive method, both AE and AC latencies declined from the first half to the second half of the precriterion trials, with the AE latency declining more. The AE decline was interpreted as reflecting shrinkage in the hypothesis pool with experience—as hypotheses are tried and rejected, some are retained in short-term memory. With the simultaneous method, however, AC latencies declined but AE latencies did not. One interpretation of this result is that the hypothesis pool did not shrink under simultaneous conditions—or the memory load is such that rejected hypotheses are *not* retained in short-term memory.

Taken as a whole, these data seem to indicate that increasing the memory load in a concept-identification task has the effect of increasing the likelihood that *S*s actually will sample with replacement from a pool of hypotheses made up of the allowable response assignments.

REFERENCES

Bower, G., & Trabasso, T. Reversals prior to solution in concept identification. *J. exp. Psychol.*, 1963, **66**, 409–418.

Bower, G., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964. Pp. 32–96.

Erickson, J. R., Zajkowski, M. M., & Ehmann, E. D. All-or-none assumptions in concept identification: Analysis of latency data. *J. exp. Psychol.*, 1966, **72**, 690–697.

Holstein, S. R., & Premack, D. On the different effects of random reinforcement and presolution reversal on human concept identification. *J. exp. Psychol.*, 1965, **70**, 335–337.

Levine, M. Cue neutralization: The effects of random reinforcements upon discrimination learning. *J. exp. Psychol.*, 1962, **63**, 438–443.

Millward, R. Latency in a modified paired-associate learning experiment. *J. verbal Learn. verbal Behav.*, 1964, **3**, 309–316.

Restle, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

---

[3] The increases in errors and in latency discussed here are also found when individual problems are examined.

RESTLE, F., & EMMERICH, O. Memory in concept attainment: Effects of giving several problems concurrently. *J. exp. Psychol.,* 1966, **71,** 794–799.

SUPPES, P., & GINSBERG, R. A fundamental property of all-or-none models, binominal distribution of responses prior to conditioning, with application to concept formation in children. *Psychol. Rev.,* 1963, **70,** 139–161.

TRABASSO, T., & BOWER, G. Memory in concept identification. *Psychon. Sci.,* 1964, **1,** 133–134. (a)

TRABASSO, T., & BOWER, G. Presolution reversal and dimension shifts in concept identification. *J. exp. Psychol.,* 1964, **67,** 398–399. (b)

TRABASSO, T., & BOWER, G. Presolution dimensional shifts in concept identification: A test of the cue-sampling axiom. *J. math. Psychol.,* 1966, **3,** 163–173.

(Received June 23, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 219–224

EFFECTS OF PARTIAL REINFORCEMENT ON DISCRIMINATION LEARNING AND SUBSEQUENT REVERSAL OR EXTINCTION[1]

ELLEN S. PENNES[2] AND JAMES R. ISON

University of Rochester

47 rats received 40 differential instrumental conditioning trials interspersed with 20 choice trials, with either 100% (CR) or 50% (PR) reinforcement on the positive stimulus and 0% reinforcement on the negative stimulus. This was followed by either 30 reversal or 15 extinction trials. In initial acquisition speed to the positive stimulus was little affected by PR whereas speed to the negative stimulus was greater under PR. Thus, the retardation of choice behavior observed with PR seems attributable to its attenuating effect on the inhibitory process associated with the incorrect response rather than an effect on the excitatory process associated with the correct response. In either extinction or reversal PR led to increased resistance to extinction. Additionally the approach to the formerly correct stimulus was of greater strength in reversal than in extinction which was ascribed to a positive induction effect.

It has been routinely demonstrated that the learning of a simultaneous discrimination is retarded if the correct response is only partially reinforced (cf. Jenkins & Stanley, 1950; Lewis, 1960). However, there seems to be no analytic study of the behavioral processes responsible for this common finding, perhaps because it can be ascribed simply to the debilitating effect of partial reinforcement (PR) on the strength of the approach to the positive stimulus. There is no direct evidence supporting this simple hypothesis; the simultaneous discrimination problem in which the effect is demonstrable confounds the development of approach strength to the positive stimulus with the concurrent development of avoidance of the negative stimulus. With this procedure it is equally logical that retardation results because PR of the correct response leads to the maintenance of approach behavior to the incorrect stimulus.

This latter hypothesis is additionally suggested by the cogent analysis of within-*S* PR effects provided by Amsel and his coworkers (Amsel, Rashotte, & MacKinnon, 1966). Following the earlier analyses of between-*S* effects (Amsel, 1958), enhanced resistance to extinction (PRE) resulting from PR is attributed to the counterconditioning of an instrumental approach response to the stimuli of anticipatory frustration during acquisition. Thus, during subsequent extinction, the probability of avoidance elicitation is reduced in a partially reinforced group relative to a continuously reinforced group. The critical process is the conditioning of approach behavior to the cues of anticipatory frustration with the result that future occurrences of anticipatory frustration should elicit the same approach

---

[1] This research was supported by National Science Foundation Research Grants GB 1505 and GB 4373 to James R. Ison, and performed in part during the senior author's tenure as a National Science Foundation undergraduate research participant.

[2] Now at the University of Michigan.

behavior. It may be anticipated, therefore, that the PRE will occur on a continuously rewarded S-R combination if the organism has had prior partial reinforcement training on a second S-R combination. In support of this Ross (1964) found a generalized PRE from a partially reinforced S-R combination to a continuously reinforced S-R combination, provided that the responses were topographically similar.

This analysis, with the additional assumption that the development of inhibition to a never reinforced stimulus in discrimination learning follows the same laws as the development of inhibition to a previously reinforced stimulus in simple instrumental extinction, indicates that PR of the correct response will maintain the strength of the incorrect response in discrimination learning. In effect, it is anticipated that the concurrent administration of two reinforcement schedules to topographically similar responses in discrimination learning is equivalent to their sequential administration to similar responses in instrumental conditioning as used by Ross (1964).

The method of single stimulus presentation is well suited to the exploration of this problem because it permits assessment of the separate response strengths to each of the discriminanda. In the present experiment this method was supplemented by the simultaneous presentation of both stimuli on one-third of the trials in order to provide additional data on choice behavior. After acquisition *S*s received either extinction or reversal training.

METHOD

Subjects.—The *S*s were 48 male albino rats, approximately 90 days old, obtained from The Holtzman Company, Madison, Wisconsin. One *S* died during the course of the experiment.

Apparatus.—The differential conditioning box consisted of six parallel alleyways, each 36 in. long, 3½ in. wide, and 4 in. high, to any of which a start box, 12 in. long, 3½ in. wide, and 4 in. high, could be attached. The start box and its side-action retrace door were painted midgray. Three alleys were painted a flat off-white and three flat black. These were arranged so that two presented a choice with white on left, black on right; two a choice with the stimuli reversed; and two were presented singly for differential conditioning. Appropriately painted wooden baffles at the end of the alleyways concealed the presence or absence of a drilled aluminum block which held .20 cc. of the liquid reinforcer—15% sucrose solution. Photocells, attached to the start door and 12 in. and 24 in. down the alleys, associated with two Hunter clocks, recorded start and running times.

Procedure.—The *S*s were maintained on a 23-hr. water deprivation schedule (food ad lib) with water available for 1 hr. immediately after the daily treatment. They were placed on this schedule and were individually handled for 4 min. each day for 5 days prior to runway training. On the last 2 such days they were permitted to roam the maze in groups of three for 5 min.

On each day of acquisition four forced and two choice trials were administered in the pattern CFFFCF. Trials 2 and 6 were forced to the opposite of that previously chosen whereas the order of stimuli for Trials 3 and 4, as well as the position of choice points (i.e., white or black on left) on Trials 1 and 5, was counterbalanced to equate responses to each color. Stage I (acquisition) was continued for 10 days (30 trials to each stimulus) with white positive for all *S*s. Half of the *S*s received 30 reinforcements on white (CR) whereas half received but 15 (PR). In the latter group the order of reinforcements was random with the constraints that at least one occurred each day, and no more than three occurred in any two consecutive days. In this stage black was never reinforced.

In Stage II half of each group was randomly assigned to reversal learning and half to extinction. Reversal followed the same trial sequence as acquisition, except that black was associated with 100% reinforcement for all *S*s and white with 0%. Extinction was run at three trials per day to the single white stimulus. The second stage continued for 5 days.

On all nonreinforced trials the food cup was absent and *S* was detained in the arm for 8 sec. The minimum intertrial interval was 10 min. in the first stage, 5 min. in the second stage.

RESULTS

Choice data in acquisition and reversal is presented in Fig. 1. In agreement with past results the partially reinforced animals approached the correct alternative less frequently in acquisition and in subsequent reversal changed less rapidly in their choice of the formerly correct brightness. The reversal stage was not sufficiently extensive to detect whether differences would appear similarly in trials to a criterion of reversal performance. A simple analysis of variance of correct responses in acquisition yielded a significant effect of reinforcement schedule, $F\ (1,\ 42) = 57.3,\ p < .01$. A mixed-design analysis of variance of the reversal data was performed in which the within-*S* variable was performance in the first two reversal days compared to the last two. This yielded a significant Schedule × Days interaction, $F\ (1, 22) = 6.0,\ p < .05$.

Start and running speeds to each stimulus on Trials 3 and 4 of each day of acquisition are presented in Fig. 2.

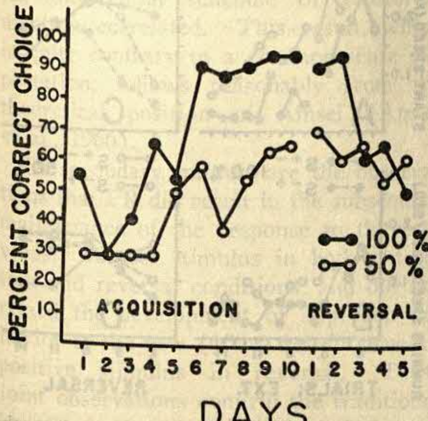


FIG. 1. Percent choice of *S*+ during acquisition and reversal.

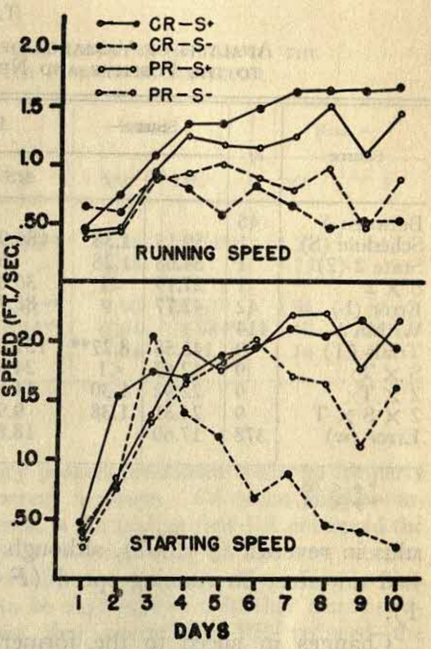


FIG. 2. Starting and running speeds to the S+ and S− in acquisition.

Mixed-design analyses of variance[3] of these data are presented in Table 1. The reinforcement schedule on the correct stimulus had no effect on start speed, whereas it had the major effect of maintaining speed of response to the incorrect stimulus ($p < .01$). In Fig. 3 are presented start and running speeds to the formerly positive stimulus in the extinction condition (A and B) and to each stimulus in the reversal condition (C and D). Analyses of these data are presented in Tables 2 and 3. First, it is clear that in each condition PR on the correct response in acquisition maintained the strength of that response during subsequent nonreinforcement. In addition PR in acquisition resulted in a subsequent faster starting speed to the new positive stim-

---

[3] In these analyses scores were estimated at the cell mean for the missing *S*, and *df* were adjusted accordingly.

TABLE 1

ANALYSIS OF VARIANCE OF STARTING AND RUNNING SPEEDS TO THE POSITIVE AND NEGATIVE STIMULI IN ACQUISITION

Source	*df*	Start +		Run +		Start −		Run −	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*	45								
Schedule (S)	1	59.15	1.38	186.95	2.25	2712.40	14.78**	1693.06	6.39*
State 2 (2)	1	54.68	1.28	.66	<1	91.06	<1	328.92	1.24
S × 2	1	27.19	<1	36.81	<1	26.63	<1	54.03	<1
Error (b)	42	42.77		82.78		183.81		264.91	
Within *S*	414								
Trials (T)	9	145.58	8.27**	151.63	5.08**	339.23	3.06**	183.78	1.33
S × T	9	12.67	<1	24.95	1.32	303.39	2.74**	354.08	2.56**
2 × T	9	22.92	1.30	14.88	<1	94.11	<1	191.69	1.39
2 × S × T	9	24.31	1.38	9.90	<1	40.77	<1	88.38	<1
Error (w)	378	17.60		18.86		110.98		138.16	

* $p < .05$.
** $p < .01$.

ulus in reversal ($p < .01$), although it had no effect on running speed ($F < 1$).

Changes in speed to the formerly positive stimulus in the extinction condition were typical of extinction data in simple instrumental conditioning, whereas changes in speed to the formerly positive stimulus in the reversal condition were atypical in that there was little decline over the first several days following CR and even a slight increase in speed following PR.

No attempt was made to relate differences in group speed scores in the differential conditioning trials to group choice scores. In general individual *S*s responded with nondifferential speeds to both stimuli for an initial sequence of trials followed by a rapid transition to differential speeds. In all but one *S* differentiation was observed in speed scores either prior to or concurrent with differentiation in choice. Retardation in choice, where it occurred, resulted because of a continued position response.

TABLE 2

ANALYSES OF VARIANCE OF STARTING AND RUNNING SPEEDS TO THE FORMERLY POSITIVE STIMULUS IN EXTINCTION

Source	*df*	Start +		Run +	
		MS	*F*	*MS*	*F*
Between	22				
Schedule (S)	1	35.45	5.99*	.10	<1
Error (b)	21	5.92		.84	
Within (S)	92				
Days (D)	4	9.06	18.13**	2.33	25.78**
D × S	4	1.58	3.16*	.35	3.89**
Error (w)	84	.50		.09	

* $p < .05$.
** $p < .01$.

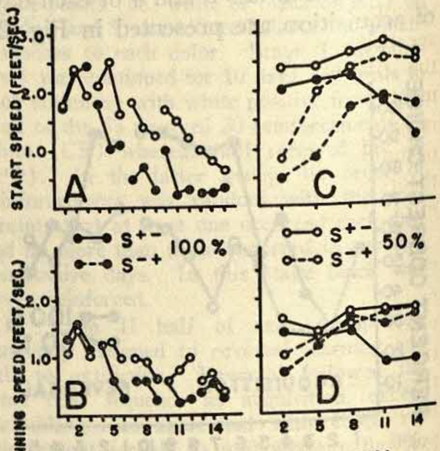


FIG. 3. Starting and running speeds to the former S+ in extinction (A and B) and to the former S+ and S− in reversal (C and D).

TABLE 3

ANALYSES OF VARIANCE OF STARTING AND RUNNING SPEEDS TO THE FORMERLY POSITIVE AND NEGATIVE STIMULI IN REVERSAL

Source	*df*	Start +		Run +		Start −		Run −	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*	23								
Schedule (S)	1	15.66	3.96	7.07	9.15**	15.40	9.74**	.27	<1
Error (b)	22	3.95		.77		1.58		.73	
Within *S*	96								
Trials (T)	4	1.15	2.71*	.42	6.21**	9.90	18.67	2.94	18.37
T × S	4	1.45	3.41*	.72	10.54**	2.64	4.98**	.10	<1
Error (w)	88	.42		.07		.53		.16	

* $p < .05$.
** $p < .01$.

DISCUSSION

The result of primary interest is that the reinforcement schedule on the correct response had but little differential effect on the speed of that response, but had a major effect of maintaining response speed to the incorrect stimulus. If it is assumed that these speeds in differential conditioning reflect the strengths of excitatory and inhibitory processes responsible for choice behavior, then it follows that the effect of PR on simultaneous discrimination performance is primarily localized in processes associated with the nonreinforced incorrect stimulus, and not in those processes associated with the stimulus with which the differential schedule of reinforcement is correlated. This result, which is quite contrary to a common-sense expectation, follows reasonably from the theoretical position of Amsel (Amsel et al., 1966).

Of secondary interest are the observations that PR did result in the subsequent maintenance of the response to the previously correct stimulus in both extinction and reversal conditions, and did not retard the development of approach behavior to the new (100% reinforcement) positive stimulus in reversal. These joint observations confirm the traditional hypothesis that retardation in choice reversal following PR is localized in its delaying effect on the developing inhibitory process associated with the formerly correct response. Of some further interest is the finding that PR enhanced the speed of responding to the new correct stimulus in reversal. Most simply this can be explained as a further demonstration that acquisition PR reduced the amount of inhibition associated with the incorrect stimulus, a reduction that was most pronounced on the start measure.

Finally, it was observed that extinction of the former correct response was much delayed in the reversal condition compared to the extinction condition. This might be attributable to the generalized effects of reinforcement on the other stimulus adding to excitatory strength in reversal. However, a similar effect was demonstrated by Miller and Senf (in press) in which presentation of the formerly reinforced stimulus was alternated with presentations of the original nonreinforced stimulus which *remained* nonreinforced. Thus, the continued maintenance of response strength in their experiment obviously cannot be attributed to a generalized effect of reinforcement on the other stimulus and, instead, provided support for the operation of the principle of positive induction (Pavlov, 1927) in differential instrumental conditioning. This induction process clearly should be operative in the present experiment and would account for the obtained difference between the reversal and extinction conditions.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A., RASHOTTE, M. E., & MACKINNON, J. R. Partial reinforcement effects within-subject and between-subjects. *Psychol. Monogr.*, 1966, **80** (20, Whole No. 628).

JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

SENF, G. M., & MILLER, N. E. Evidence for positive induction in discrimination learning. *J. comp. physiol. Psychol.*, in press.

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

ROSS, R. R. Positive and negative partial reinforcement extinction effects carried through continuous reinforcement, changed motivation, and changed response. *J. exp. Psychol.*, 1964, **68**, 492–502.

(Received June 23, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 225–229

EYE-MOVEMENT PARAMETERS IN PATTERN RECOGNITION

JOHN D. GOULD AND AMY SCHAFFER

IBM Research Center, Yorktown Heights, New York

In visually scanning 9 simultaneously presented patterns, human Ss fixated significantly longer (a) on patterns they were looking for (i.e., on targets) than on patterns they were not looking for; and (b) on patterns that exactly matched a memorized standard pattern than on patterns that differed from this standard. These results suggest that (a) Ss may have a generalized tendency, within certain limits, to look longer at stimuli they are looking for than at stimuli they are not looking for; and (b) Ss make definite analytic comparisons of the details and differences between a memorized standard pattern and a given fixated pattern, as opposed to detecting a fixated pattern in a more immediate or holistic manner.

Since little or no perception occurs during the saccadic movement of the eyes between points of fixation, the parameters of eye movements intrinsically involved in visual perception are number, duration, and location of eye fixations. The main purposes of an on-going series of experiments, in which the correlations of these eye-movement parameters with stimulus parameters are analyzed, are to determine these relationships and to ascertain their implications for human visual perception.

The present study examines more closely a finding of two previous experiments (Gould & Schaffer, 1965c; Gould & Schaffer, 1966) in which human *S*s were presented with a stimulus containing nine patterns of asterisks, arranged in a 3 × 3 matrix. An *S* initially fixated the center, or standard, pattern, which defined the target pattern for the trial, and then determined how many of the eight peripheral patterns were identical to that pattern in the center cell, i.e., how many target patterns were present. Eye-movement records indicated that, although mean number of fixations on target patterns did not differ from mean number of fixations on nontarget patterns, *S*s fixated target patterns significantly longer than they fixated nontarget patterns.

Two hypotheses would seem to account for the longer fixations on target patterns. First, perhaps longer fixations are needed to detect an exact match between the memorized standard pattern and a given fixated pattern than to detect a difference between the two. Alternatively, it may be that *S*s fixate longer on patterns they are looking for than on patterns they are not looking for, regardless of the specific characteristics of each pattern.

To test between these alternatives, one group of *S*s, Match Group (MG), was instructed, as in previous studies, to look for patterns which exactly matched the center standard (i.e., a target pattern was a matching pattern). A second group of *S*s, No Match Group (NMG), was instructed to look for patterns which differed from the center pattern (i.e., a target pattern was a nonmatching pattern).

METHOD

In each of the two groups were nine paid volunteers from a nearby college. Each stimulus contained nine patterns of asterisks arranged in a 3 × 3 matrix. Upon projection of the stimulus, each *S* was first required to fixate the center pattern of the matrix, since this pattern defined the target pattern, and then to scan the eight peripheral matrix positions to determine the number (or frequency), varied two through six, of target patterns. The pattern in the center cell was not considered one of the targets. Target frequency was varied both to determine its effect upon eye-movement parameters and to ascertain the accuracy of *S*'s perception. Target frequencies of two through six were selected because, among other reasons, on the average an equal number of targets and nontargets was presented, and thus any difference between the variance estimate of a given eye-movement parameter on target patterns and that estimate on nontarget patterns would not be affected by sample size.

On each of 3 days, each *S* scanned three matrices at each of the five target frequency conditions. Time spent scanning and number of errors were recorded. On the last 2 of these days, eye movements were also recorded by means of a closed circuit television, corneal-reflectance, eye-marker system which yielded filmed records (18 frames/sec) of the stimulus field and the position of the fixation point of *S*'s right eye within it. This apparatus was originally developed by Mackworth and Mackworth (1958); an extensive discussion of our modifications and recording procedures was given in Gould and Schaffer (1965b). Data were transduced to punched cards for computer analysis.

Prior to the start of each trial, each *S* was required to fixate a point about 3° from the center pattern. This permitted a check of the correspondence between the eye-marker spot and the point in the visual field at which *S* was looking, since *E* could view both on the television monitor; it also meant that the fixation position of each *S* was always the same at stimulus onset. Each *S* was instructed to scan each pattern as rapidly as possible without sacrificing accuracy. Correct answer feedback was provided after each trial.

TABLE 1

MEAN DURATION (SEC.) OF FIXATIONS

	Targets	Nontargets	
Match	.47 sec. *S*s = MG	.44 sec. *S*s = NMG	.455 sec.
No Match	.46 sec. *S*s = NMG	.35 sec. *S*s = MG	.405 sec.
	.465 sec.	.395 sec.	

Each stimulus pattern and its random position in the 3×3 stimulus matrix were generated by computer program.[1] Each pattern consisted of 10 asterisks inside a 4×4 matrix; the bottom row was always filled and an asterisk appeared within a column only if the row below it had one. Otherwise, patterns which differed from the center standard pattern did so randomly. These rules allowed the generation of 44 different patterns. Patterns subtended approximately 1.5°, with their midpoints separated by approximately 6.5°. Measured luminance of each asterisk was 16.07 ftl., and the patterns appeared against a darker background (.33 ftl.).

RESULTS

Analysis of variance of the mean duration of fixations,[2] summarized in Table 1, showed that, as hypothesized, *S*s did fixate significantly longer, $F\ (1, 16) = 54.63;\ p < .001$, in perceiving a pattern that exactly matched the memorized center pattern (.455 sec.) than in perceiving a pattern that differed from this standard (.405 sec.). Furthermore, in conformity with the second hypothesis, *S*s also fixated significantly longer, $F\ (1, 16) = 114.85;\ p < .001$, when perceiving targets (.465 sec.), i.e., what they were looking for, than when perceiving nontargets (.395 sec.), i.e., what they were not looking for. Both of these significant differences were due mainly to the much shorter mean duration of fixations on nonmatching, nontargets (.35 sec.). This mean was significantly shorter (Duncan range test; $p < .01$) than the mean durations of fixations on the other three types of patterns. None of the latter three means differed significantly from one another.

---

[1] Cyril Alberga wrote this program.

[2] Mean duration of fixations on each type of pattern was determined by a computer program that selected, for each trial, all fixations on each type of pattern and determined the mean of these durations for each type of pattern. Thus, each mean presented in the table is based upon the 135 means of the last day (9 *S*s × 5 Frequencies × 3 Trials).

Thus the effects of these two variables (Target vs. Nontarget and Match vs. No Match) seemed to be of about equal importance; however, they were not linearly additive. That is, although confounded with the random variable of *S*s, there was a tendency toward a significant Match-No Match × Target-Nontarget interaction, $F\ (1,\ 16) = 3.63;\ p < .10$. In combination (MG), these variables led to a significant difference; in opposition (NMG), they did not.

In contrast to the results for mean duration of fixations, the mean number of fixations (Table 2) on targets did not differ significantly from the mean number of fixations on nontargets ($p > .10$), and the mean number of fixations on matches did not differ significantly from the mean number of fixations on nonmatches ($p > .10$). However, their interaction was significant ($p < .001$). A Duncan range test showed that the four means fell into two classes: (Nonmatching, Nontargets [4.19 fixations] and Matching, Targets [4.48]) and (Matching, Targets [4.48], Matching, Nontargets [4.96], and Nonmatching, Targets [5.00]). Means within the same class, i.e., within the same pair of parentheses, do not differ significantly from each other; means which are not in the same class do differ from each other at the .01 significance level. It was also found that NMG *S*s refixated the center standard pattern significantly more often (1.71 refixations) than did MG *S*s (.60 refixation), $F\ (1,\ 16) = 38.36;\ p < .001$, possibly suggesting that the former group tended to forget the standard pattern more often than the latter group.

The *S*s reported erroneously on about 10% of the trials; there was no significant difference between the two groups in number of errors.

TABLE 2

MEAN NUMBER OF FIXATIONS

	Targets	Nontargets	
Match	4.48 *S*s = MG	4.96 *S*s = NMG	4.72
No Match	5.00 *S*s = NMG	4.19 *S*s = MG	4.60
	4.74	4.58	4.68

No eye-movement parameter showed any simple, significant relation to position of patterns on the matrix, trials, order of perceiving patterns, or frequency of targets (when the latter was corrected for number of targets present). This last result is consistent with that found using targets defined on the basis of the arithmetic sum of three digits (Gould & Schaffer, 1965a), but differs from that found when single digits were used as targets (Gould & Schaffer, 1965b; Schaffer & Gould, 1964). In the latter case a general increase in the total mean number of fixations was observed with an increase in the number of target digits on the stimulus. Although a general conclusion from these four studies might be that the total mean number of fixations on a stimulus increases with the number of targets on that stimulus only for stimuli with quite simple targets and nontargets, there are enough other differences among these studies to render this conclusion questionable.

DISCUSSION

The two main findings of this study support both of the original hypotheses: *S*s tend to look longer at visual patterns they are looking for than at visual patterns they are not looking for, and they also look longer at patterns which exactly match a standard pattern than at those that differ from this standard.

It is tempting to generalize the finding of longer fixations on target patterns to other types of stimuli. While the limits of this generalization are not yet known, limits do exist, since no difference was found between the duration of fixations on targets and the duration of fixations on nontargets for targets and nontargets defined by the arithmetic sums of three digits (Gould & Schaffer, 1965a).

It might be hypothesized that the longer duration of fixations on target patterns was at least partially due to the process of counting targets. However, evidence does not support this hypothesis, since in the study just cited *S*s were also required to count targets and their duration of fixations on targets was not longer than their duration of fixations on nontargets.

In a considerably different task, Neisser (1963) showed that when *S*s scan a list of items for the absence of a target character the mean time to perceive the items containing this character (i.e., nontarget items) significantly exceeds the mean time, obtained in a separate condition, to perceive the absence of a target character (i.e., nontarget items) when *S*s scan for its presence. This is in agreement with the two nontarget means of Table 1. Thus, in both studies *S*s looked longer at nontarget stimuli which matched a memorized standard than at those that differed from this standard.

The second main finding of the present study suggests that *S*s made definite analytic comparisons of the details and differences between a standard pattern and a given fixated pattern, as opposed to detecting a fixated pattern in a more immediate or holistic way. If *S*s do make these comparisons, then the fact that more time is taken to process a pattern which matches a memorized standard than one that differs from this standard is, of course, not surprising. In perceiving a match, a "complete" set of comparisons is required, whereas in perceiving a nonmatch, a "complete" set of comparisons is typically not required, since the detection of a single difference is sufficient. If, on the other hand, patterns are detected in a more immediate or holistic way, the explanation of longer fixations on patterns that match a memorized standard pattern than on patterns that differ from this standard would not be obvious. These differential fixation durations also suggest that these comparisons are not carried out simultaneously or entirely in parallel.

With respect to the number of fixations, the mean number of fixations (9.39) on peripheral patterns was similar to the actual number of peripheral patterns (8), i.e., although there were individual differences among *S*s, generally each peripheral pattern was fixated only once. Mean number of refixations of the center pattern was also low (1.15 refixations). The significance of these findings is that once *S* initially fixates the center pattern he presumably memorizes it, since he refixates it only once while he is in the process of determining its relationship to eight other patterns. Perhaps his subsequent identification of patterns which either exactly match this standard pattern or are very similar to it aids this memory process, or perhaps, through peripheral vision, he continues to perceive part of the center pattern while he centrally fixates one of the other patterns. The *S*s do not look back and forth between the center pattern and a given peripheral pattern in determining whether they are identical or different. It should be noted that the stimulus patterns can hardly be considered simple; the difference(s) between any two of the 44 different possible patterns was sometimes quite subtle.

REFERENCES

Gould, J. D., & Schaffer, A. Eye-movement patterns during visual information processing. *Psychon. Sci.*, 1965, **3**, 317-318. (a)

Gould, J. D., & Schaffer, A. Eye-movement patterns in scanning numeric displays. *Percept. mot. Skills*, 1965, **20**, 521-535. (b)

Gould, J. D., & Schaffer, A. Systematic analysis of eye movements in pattern perception. Paper presented at Eastern Psychological Association, Atlantic City, April 1965. (c)

GOULD, J. D., & SCHAFFER, A. Pattern perception and eye-movement parameters. Paper presented at Midwestern Psychological Association, Chicago, May 1966.

MACKWORTH, J. F., & MACKWORTH, N. H. Eye fixations recorded on changing visual scenes by the television eye-marker. *J. Opt. Soc. America,* 1958, **48,** 439–445.

NEISSER, U. Decision without reaction time: experiments in visual scanning. *Amer. J. Psychol.,* 1963, **76,** 376–386.

SCHAFFER, A., & GOULD, J. D. Eye-movement patterns as a function of previous tachistoscopic practice. *Percept. mot. Skills,* 1964, **19,** 701–702.

(Received June 24, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 230–235

EFFECTS OF COMPOSITION OF THE POSITIVE CATEGORY ON CONCEPT LEARNING[1]

ROBERT C. HAYGOOD AND JAMES V. DEVINE
Kansas State University

2 experiments were performed to clarify the reason that positive instances are superior to negative instances in conjunctive concept learning. In conjunction, all positive instances embody both relevant attributes (e.g., red and square). In nonconjunctive concepts, the positive category includes other kinds of instances (e.g., inclusive disjunction: red or square or both). Using nonconjunctive rules, increasing the proportion of instances embodying both relevant attributes improved performance with the proportion of positive instances held constant. Changing the proportion of positive instances had little effect when the composition of the positive category was held constant, confirming that it is instances with both relevant attributes, rather than positive instances as such, that influence S's performance.

Hovland (1952) presented a detailed analysis of conceptual problems, in which he demonstrated that positive instances (examples of the concept) in general transmit more information about a concept than negative instances (nonexamples). Thus it ordinarily takes more negatives to present complete information necessary to solve a concept-learning problem, an effect which is magnified as the number of irrelevant dimensions and values on dimensions increases. Hovland suggested that superior performance of *S*s with positive instances might result in large part from the informational superiority of such instances.

However, Hovland and Weiss (1953) subsequently found a clear superiority of all-positive sequences over mixed and pure-negative sequences when all sequences had been equated for information transmission. They attributed this result to differences in difficulty of assimilating information concerning what the concept "is" as compared with assimilating information concerning what it "is not." In particular, they pointed out that in positive instances, the correct combination of characteristics defining the concept is directly perceptible, while in negative instances, the correct combination is never in direct view.

Role of conceptual rule.—In the Hovland and Weiss experiment, only two-dimensional conjunctive concepts were used, although in two of their experiments several alternative combinations of values from the two dimensions were correct. This poses two difficulties for interpretation. First, Bruner, Goodnow, and Austin (1956) have shown that powerful strategies utilizing positive focus exist for conjunction. For other rules, however, such as inclusive disjunction and conditional, the optimum strategy may be to begin with a careful scan of negative instances, or with a comparison of positive and negative instances (Hunt, 1962). Thus the demonstrated superiority of positive instances might be primarily an artifact of the strategies available for conjunctive problems.

Secondly, in conjunctive problems,

---

[1] This research was supported in part by a research grant MH 11283 from the United States Public Health Service, National Institute of Mental Health.

TABLE 1

DISTRIBUTION OF INSTANCES INTO POSITIVE AND NEGATIVE CATEGORIES FOR FOUR BASIC CONCEPTUAL RULES

Stimulus Patterns		Logical Truth Table	Conceptual Rules			
Color	Form	Red Square	Conjunction	Inclusive Disjunction	Conditional	Biconditional
Red	Square	T T	+	+	+	+
Red	Not Square	T F	−	+	−	−
Not Red	Square	F T	−	+	+	−
Not Red	Not Square	F F	−	−	+	+

all members of the positive category exhibit both of the relevant characteristics; these instances may be called TT instances.[2] When conjunctive concepts are used, increasing the proportion of positive instances [p(+)] necessarily increases the proportion of TT instances [p (TT)]. The results of Hovland and Weiss are thus ambiguous, since it is not clear how much of the beneficial effect of increasing p (+) actually results from the necessary concomitant variation of p (TT) in conjunctive problems.

Table 1 shows the distribution of instances into positive and negative categories for the four basic two-dimensional conceptual rules. In nonconjunctive rules, some positive instances may fail to have one or the other relevant characteristic (inclusive disjunction) or may fail to have both (biconditional). In such rules, it is possible to vary p (+) and p (TT) independently. This permits research to be designed that will not only resolve the ambiguity of Hovland and Weiss's finding, but also determine if their results can be generalized to other conceptual rules.

Role of instructions and methodology.—The experimental task used by Hovland and Weiss was of the type Haygood and Bourne (1965) have termed "attribute identification" (AI). In AI, the conceptual rule is explained to *S*, and his task is to discover the relevant attributes. Two other basic types of experimental task may be created through the use of instructions. These are (*a*) rule learning (RL), in which *S* is told the relevant attributes and must discover the correct rule, and (*b*) complete learning (CL), in which *S* is told neither the attributes nor the rule. The generalizability of the earlier findings to RL and CL instructional conditions is unknown, and one of the purposes of the present study was to explore this question.

The method employed by Hovland and Weiss was to present the entire stimulus sequence, either simultaneously or sequentially, to a passive *S*. At the end of the presentation, *S* was asked to show his attainment of the concept by defining or naming it. The measure of performance used was simply the number of *S*s attaining the concept. It seems desirable to extend the study of this problem to the more complex discrimination-learning paradigm presently used in concept-identification studies, using the more sensitive performance measures of errors and trials to criterion.

---

[2] Selection of two attributes as relevant in a concept-learning problem generates a contingency table consisting of the four categories of stimulus patterns defined by the presence or absence of the two attributes. For example, given red and square as attributes, the four contingencies are red square, red not-square, not-red square, and not-red not-square. This contingency table is equivalent to a truth table in symbolic logic, in which the four contingencies are designated TT, TF, FT, and FF (T standing for true and F for false). The specification of a logical rule further maps these four contingencies onto a two-response system of examples and nonexamples of the concept. For example, in a conjunctive concept, the first contingency (TT, or red square) is placed in the positive category while the other three are negative. The four basic two-dimensional rules are shown in Table 1; the plus and minus signs indicate that the instance is positive or negative, respectively.

General Procedure and Apparatus

The task was essentially the same as that described by Haygood and Bourne (1965), in that *S* was required to learn to sort a series of stimulus patterns into two categories, examples and nonexamples of a concept. The stimuli consisted of geometric designs which varied along four ternary dimensions: Color (red, yellow, or blue), Size (large, medium, or small), Shape (triangle, square, or hexagon), and Number (1, 2, or 3 fig. in each pattern). The correct categorization of each pattern was determined both by the pair of attributes relevant in the problem and by the conceptual rule being employed. For example, with red and square as relevant attributes and inclusive disjunction as the rule, all patterns that are either red or square or both would be examples of the concept. Similarly, with biconditional as the rule, patterns that are both red and square *or* neither red nor square would be examples (see Table 1).

Each *S* was given detailed instructions at the outset which described the stimulus population, the nature of a concept, the method of responding, the meaning of the feedback signal lamps, and the criterion of problem solution. In addition, each *S* was given the special instructions appropriate to his instruction condition. In RL, *S*s were told the pair of relevant attributes, and were given a special card naming the attributes. In AI, the correct rule was carefully explained to *S*, and a card showing a Venn diagram of the rule was provided. The *S* was required to show his facility with the rule by applying it to examples. In CL, neither the rule nor the attributes were given to *S*, but the exclusion rule (red *but not* square) was used in explaining the nature of a two-dimensional concept. A sample card showing the levels of the various dimensions was left exposed for *S*'s reference at all times.

The stimulus patterns were rear projected on a screen directly in front of *S*. To each pattern, *S* responded by pushing one of two buttons, labeled YES and NO, on a control panel below the screen. After a .5 sec. delay, one of two feedback signal lamps, located directly above the response buttons, was turned on for 1.0 sec. to indicate to *S* what the correct response was. Following information feedback, a period of 3.5 sec. intervened before the presentation of the next stimulus pattern.

Performance was self-paced, in that *S* was allowed as much time as needed to make his response, and instructions stressed accuracy rather than speed. In most problems, the criterion of problem solution was 16 consecutive correct categorizations. Under some conditions, certain critical instances were infrequent, and the criterion run was lengthened to 25 or 32 to insure that these instances appeared during the run.

In addition to the projection screen and *S*'s control panel, the apparatus consisted of four major parts: (*a*) a Dunning Animatic 16-mm. strip film projector used to present the stimulus patterns, (*b*) an Esterline-Angus event recorder used to record both *S*'s response and the correct response on each trial, (*c*) an electronic timing unit, which controlled the delay of feedback, length of feedback, and post-feedback interval, and (*d*) a Western Union tape reader with precoded tape, used to control the sequence of information feedback.

Experiment I

The purpose of this experiment was to determine the effect of varying p (TT) within the positive category while holding the overall p (+) constant at .50. As previously noted, such independent variation of p (TT) and p (+) can occur only with nonconjunctive rules; inclusive disjunction and biconditional problems were selected to provide maximum flexibility.

Subjects.—Ninety-six students from introductory psychology classes at Kansas State University served as *S*s and received class credit for participation. All *S*s were naive in that they had not previously participated in a concept-learning experiment. The *S*s were assigned to treatment combinations in order of appearance; each *S* was given only a single problem. Six *S*s were unable to solve during the allotted time (2 hr.) and were replaced. These *S*s were largely in the most difficult conditions.

Design.—A 4 × 3 × 2 × 2 factorial design was used, with four levels of p(TT) within the positive category (.2, .4, .6, and .8), three instructional conditions (RL, AI, and CL), two conceptual rules (inclusive disjunction and biconditional), and two different pairs of relevant attributes (three-large and blue-triangle).

Results.—Mean errors to criterion as a function of p(TT) is shown in Fig. 1 for both inclusive disjunction and biconditional rules. Increasing p(TT) in the positive category clearly improved performance, *F* (3, 48) = 7.71, $p < .01$. Although the effect appears more regular for inclusive disjunction than for biconditional, the interaction was not significant. Biconditional problems were more difficult than inclusive disjunction problems, as has been consistently found in the past (Haygood & Bourne, 1965; Hunt & Kreuter, 1962; Neisser & Weene, 1962), *F* (1, 48) = 55.56, $p < .01$.

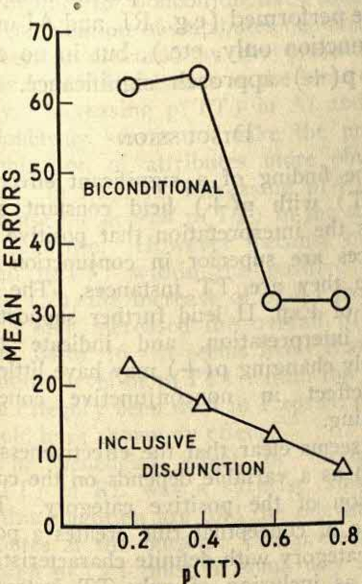


Fig. 1. Mean errors as a function of p(TT) in the positive category for two different conceptual rules, Exp. I.

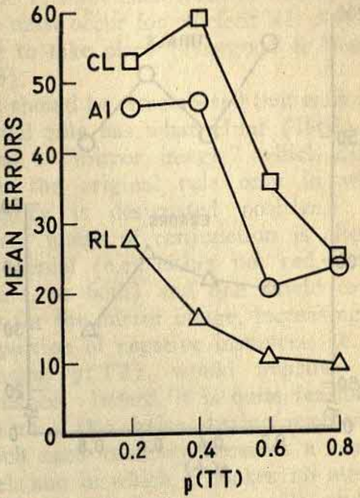


Fig. 2. Mean errors as a function of p(TT) in the positive category for three instructional conditions, Exp. I.

The same function for the three instructional conditions is shown in Fig. 2. CL was more difficult than AI, which in turn was more difficult than RL, *F* (2, 48) = 13.98, $p < .01$. The difference between CL and AI was not as marked as that found in earlier work (Haygood & Bourne, 1965). This resulted from the fact that for the blue-triangle disjunction and three-large biconditional, CL was easier than AI, giving a significant Instructions × Rules × Attributes interaction, *F* (2, 48) = 2.64, $p < .05$. These inversions occurred at .2 and .6 p(TT) for the three-large problems and .4 for blue-triangle, giving a significant p(TT) × Attributes interaction, *F* (3, 48) = 4.73, $p < .01$. It appears that these represent largely sampling error; the elimination of CL nonsolvers and replacement by better *S*s probably accounts for a part of the interactions. Figure 2 shows that the effect of p(TT) was approximately the same in RL, AI, and CL; the p (TT) × In-

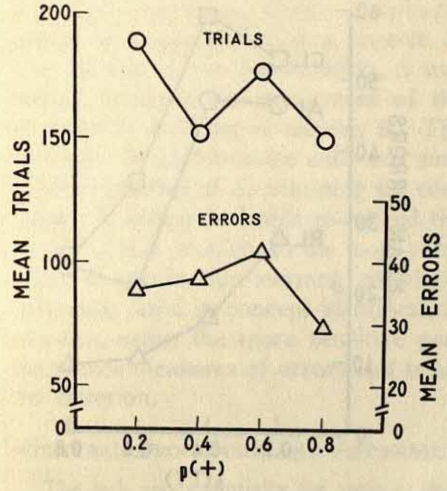


FIG. 3. Mean errors and trials to last error as a function of p(+), Exp. II.

structional Condition interaction was not significant. Pair of relevant attributes was not a significant source of variance. Analysis of trials to last error gave essentially the same results, except that the effect of p(TT) reached only the .05 level of significance.

EXPERIMENT II

The approach of this experiment was essentially the converse of Exp. I: to vary p(+) while holding p(TT) constant within the positive category. For the stimulus population used in the present study, TT instances make up .2 of the positive category for both inclusive disjunction and biconditional rules, unless the proportion is arbitrarily changed by adding or subtracting TT instances. In the absence of any rationale for modification, the natural p(TT) of .2 was used.

Subjects.—Ninety-six summer-session students at Kansas State University served as *S*s and were paid for participation. None was enrolled in introductory psychology. As in Exp. I, all *S*s were naive. Each *S* was assigned to a treatment combination in order of appearance and served for a single problem. Fifteen *S*s were unable to solve during the allotted time (2 hr.) and were replaced. All but one of these were with biconditional problems. However, they were distributed without bias with respect to p(+), with 3, 4, 5, and 3 nonsolvers in the .2 to .8 p(+) conditions.

Design.—The experimental design was identical to that of Exp. I except for the substitution of p(+) for p(TT) as the main variable of interest.

Results.—Mean errors and mean trials to last error as a function of p(+) are shown in Fig. 3. Analysis of variance of errors to criterion showed no significant effect of p(+), $F\ (3, 48) = 1.23$. As in Exp. I, the biconditional rule was more difficult than inclusive disjunction, $F\ (1, 48) = 45.65$, $p < .01$. Similarly, the order of difficulty of instructional conditions was CL most difficult and RL easiest, $F\ (2, 48) = 7.98$, $p < .01$. Pair of relevant attributes also proved to be significant, with blue-triangle easier than three-large, $F\ (1, 48) = 8.18$, $p < .05$. None of the interactions was significant.

Analysis of trials to last error showed the same results, with the F for p(+) equal to .68. Various partial analyses were performed (e.g., RL and AI only, disjunction only, etc.), but in no case did p(+) approach significance.

DISCUSSION

The finding of a significant effect of p(TT) with p(+) held constant supports the interpretation that positive instances are superior in conjunction because they are TT instances. The results of Exp. II lend further support to this interpretation, and indicate that merely changing p(+) may have little or no effect in nonconjunctive concept learning.

It seems clear that the effectiveness of p(+) as a variable depends on the composition of the positive category. The choice of conceptual rule creates a positive category with definite characteristics, such as inclusion of only TT instances in conjunction. Manipulation of p(TT) influences performance, and when the positive category contains only TT instances, p(+) will also appear to be an important variable. When the two can be separated, however, p(TT) is seen to be the critical factor.

The reason for the superiority of TT instances appears to be much the same as that originally advanced by Hovland and Weiss (1953) for the superiority of positive instances in conjunction. In conjunction, the TT (positive) instances present little more than a perceptual task in which *S* can observe the stimuli and notice what they have in common. With negative instances, on the other hand, *S*'s task is to infer the proper combination from the fact that the two attributes never appear together—though they often appear separately in the negative category. These two tasks, "perceptual" and inferential, probably differ in difficulty for the naive *S*, although the highly trained *S* should be indifferent to the source of information. Evidence indicative of this has been provided by Freibergs and Tulving (1961), who showed that the superior learning of *S*s with positive instances largely disappeared over a series of 20 problems, leaving the two kinds of instances approximately the same in learning difficulty.

Turning to nonconjunctive concepts, the combination of attributes defining the concept is obscured by the presence of other kinds of instances in the same category. Increasing p(TT) in AI and CL undoubtedly serves to make the proper combination of attributes more obvious to *S*. In Exp. II, increasing p(+) did increase the *overall* p(TT) in the stimulus sequence. However, because of the small p(TT) within the positive category (.2), the increase of p(+) from .2 to .8 only increased the overall p(TT) from .04 to .16. It seems likely that had a higher level of p(TT) within the positive category been used in Exp. II, p(+) would have shown an effect.

The beneficial effect of increasing p(TT) in RL was a surprise, since *S*s in this condition are told the relevant attributes at the outset. The presence of a high p(TT) must in some way facilitate the reductive coding of the stimulus population to the four contingencies of Table 1; some such form of coding probably must occur for efficient RL performance to take place (Haygood & Bourne, 1965).

It should be remembered that each conceptual rule has what Hunt (1962) has called a "mirror image," which differs from the original rule only in which category is designated positive. The mirror image of conjunction is alternative denial (e.g., either not red, or not square, or both) and one would expect that for the mirror image, increasing the proportion of negative instances, i.e., increasing p(TT), would improve performance. Indeed, it is quite feasible to present a two-choice sorting problem in which each category receives a neutral label, and in which *E* makes no attempt to distinguish between a rule and its mirror image. Under such conditions, the expression "positive instance" becomes meaningless.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

FREIBERGS, V., & TULVING, E. The effect of practice on utilization of information from positive and negative instances in concept identification. *Canad. J. Psychol.,* 1961, **15**, 101–106.

HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychol. Rev.,* 1965, **72**, 175–195.

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.,* 1952, **59**, 461–472.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.,* 1953, **45**, 165–182.

HUNT, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

HUNT, E. B., & KREUTER, J. M. *The development of decision trees in concept learning III: Learning the connectives.* Los Angeles: Western Management Sciences Institute, 1962.

NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *J. exp. Psychol.,* 1962, **64**, 640–645.

(Received June 24, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 236–240

STIMULUS GENERALIZATION OF HABITUATION OF THE GALVANIC SKIN RESPONSE[1]

CHARLES D. CORMAN[2]
Ohio State University

A generalization gradient of habituation of the OR occurred across groups when Ss were exposed to a new tone after hearing a similar one over a number of trials. The between-Ss design indicates that it is not necessary for S to be exposed to the full range of test stimuli in order that a gradient may be demonstrated.

Williams (1963) demonstrated a stimulus generalization of habituation gradient by repeatedly presenting a tone (380 cps) and then on four consecutive test trials presenting four new tones (670, 1,000, 1,400, and 1,850 cps). She then repeated the procedure. A gradient of generalization based on the similarity in pitch of the test stimuli to the habituation stimulus was found on the second series of test trials. This suggests that exposure to the range of test stimuli may be necessary before a gradient of generalization can be obtained. In addition, two procedural limitations exist in the Williams study. First, only test stimuli of higher pitch than the habituation stimulus were employed, and it is possible that the magnitude of the GSR might depend on absolute qualities of the stimulus such as pitch. Second, the Williams study used a fixed intertrial (or interstimulus) interval of only 7 sec. Intertrial intervals in GSR conditioning studies typically are at least 40 sec. in order to permit a recovery in base-level resistance before eliciting another response.

The present study is designed to investigate the phenomena of generalization of habituation using a between-*S*s design. In this way it can be determined if exposure to the range of test stimuli is required for a gradient to be demonstrated. Also, both increases and decreases of pitch are used, as well as longer and irregular intertrial intervals.

METHOD

Subjects.—The *S*s were 120 college students, both male and female, who chose to participate in the experiment as partial fulfillment of a laboratory requirement for an introductory psychology course.

Apparatus.—Stimulus duration was controlled by Tektronix equipment consisting of a wave-form generator and a power supply. Stimulus duration was measured by a Hunter klockkounter. The stimuli employed were four tones, 670, 1,000, 1,400, and 1,850 cps, all at approximately 60 db. (re .0002 dyne/cm^2). The tones had been judged by several *O*s to be equal in loudness. Each stimulus was generated by a Jackson audio frequency oscillator and delivered through Willson sound barrier ear-muff earphones. The GSR was measured by means of a Fels dermohmmeter and recorded from electrodes clamped to the palm and back of the left hand. The ¾ in. diameter zinc electrodes were contained in plastic cups and they were freshly filled for each *S* with an electrode jelly consisting of zinc sulphate, bacto-agar, and distilled water. The *E* recorded the response visually to the nearest 200 ohms from

---

[1] This study is based upon a dissertation submitted to the Ohio State University in partial fulfillment of the requirements for the PhD degree. The writer would like to express particular appreciation to D. D. Wickens for his advice and assistance. This investigation was supported by Public Health Service Grant MH-8526-02, from the National Institute of Mental Health.

[2] Now at the University of Southern California.

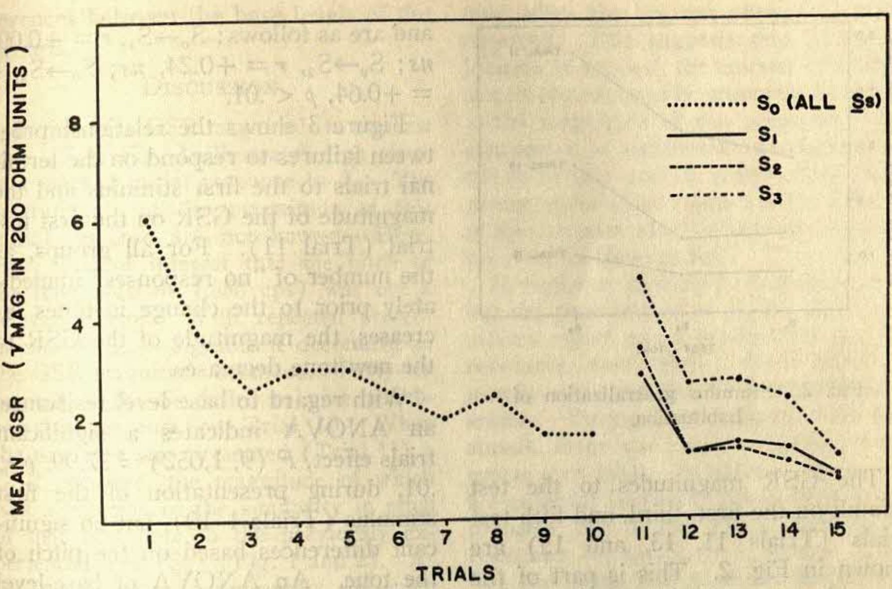


FIG. 1. Magnitude of GSRs following stimulation.

a meter on the face of the instrument and checked against a graphic record from an Esterline-Angus recorder. A reading was taken of the resistance at the time the stimulus was administered and its minimal value during any response initiated within 2 sec. after stimulation.

The *S*'s room was adjacent to the room containing the oscillators, timing equipment, dermohmmeter, graphic recorder, and *E*. The room was sound shielded and contained a small ventilating fan, which prevented *S* from hearing the operating noises produced in *E*'s room. During the experiment, *S*'s room was dimly illuminated.

Procedure.—The *S* was seated in the room and given instructions that the experiment concerned his reactions to a number of tones that would be presented at irregular intervals. The *E* then attached the electrodes to the palm and back of the *S*'s left hand, placed the earphones on *S*, and left the room. Approximately 2 min. later, the first stimulus was presented. Each *S* received 10 presentations of the habituation tone. Half of the *S*s received a tone of 670 cps and half received a tone of 1,850 cps. Each *S* then received five presentations of a new tone (670, 1,000, 1,400, or 1,850 cps). These frequencies were chosen as representing subjectively equal steps on the Mel scale of pitch (Stevens & Volkmann, 1940), and were four of the tones used by Williams. This results in test trials to S_1, S_2, or S_3, with half the *S*s receiving a higher tone than the habituation tone, half receiving a lower tone. All tones had a duration of 2,000 msec. The intertrial intervals varied between 40 and 65 sec., with a mean of 55 sec.

RESULTS

Analysis of the magnitude of the GSRs elicited by the tones was based on the square root of the magnitude (in 200 ohm units). This transformation was used to minimize the effects of a limited number of extreme scores.

GSRs to the first tone (Trials 1–10) are presented in Fig. 1. An ANOVA of these responses indicates a significant trials effect (habituation), F (9, 1,052) = 76.4, $p < .01$, and no significant difference based on the pitch of the tone.

GSRs to the test tones (Trials 11–15) are also shown in Fig. 1. An ANOVA on these responses indicates a significant trials effect, F (2, 117) = 50.71, $p < .01$, and a significant difference between groups, F (2, 117) = 5.90, $p < .01$.

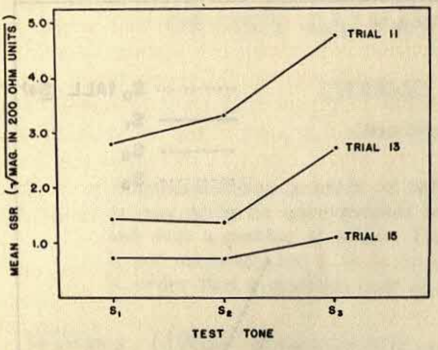


FIG. 2. Stimulus generalization of habituation.

The GSR magnitudes to the test stimuli on the first, third, and fifth test trials (Trials 11, 13, and 15) are shown in Fig. 2. This is part of the same data presented in Fig. 1, but is plotted across groups to illustrate the generalization effect.

Spearman rank-order correlation coefficients between magnitudes of GSRs on Trial 1 and Trial 11 were computed and are as follows: $S_0 \rightarrow S_1$, $r = +0.09$, *ns*; $S_0 \rightarrow S_2$, $r = +0.24$, *ns*; $S_0 \rightarrow S_3$, $r = +0.64$, $p < .01$.

Figure 3 shows the relationship between failures to respond on the terminal trials to the first stimulus and the magnitude of the GSR on the first test trial (Trial 11). For all groups, as the number of "no responses" immediately prior to the change in tones increases, the magnitude of the GSR to the new tone decreases.

With regard to base-level resistance, an ANOVA indicates a significant trials effect, $F\ (9, 1{,}052) = 27.99$, $p < .01$, during presentation of the first stimulus (Trials 1–10), but no significant differences based on the pitch of the tone. An ANOVA of base-level resistances during presentations of the test stimuli (Trials 11–15) reveals a significant trials effect, $F\ (4, 468) = 12.60$, $p < .01$, and a significant Groups × Trials interaction, $F\ (8, 468) = 2.82$, $p < .05$, but no significant dif-

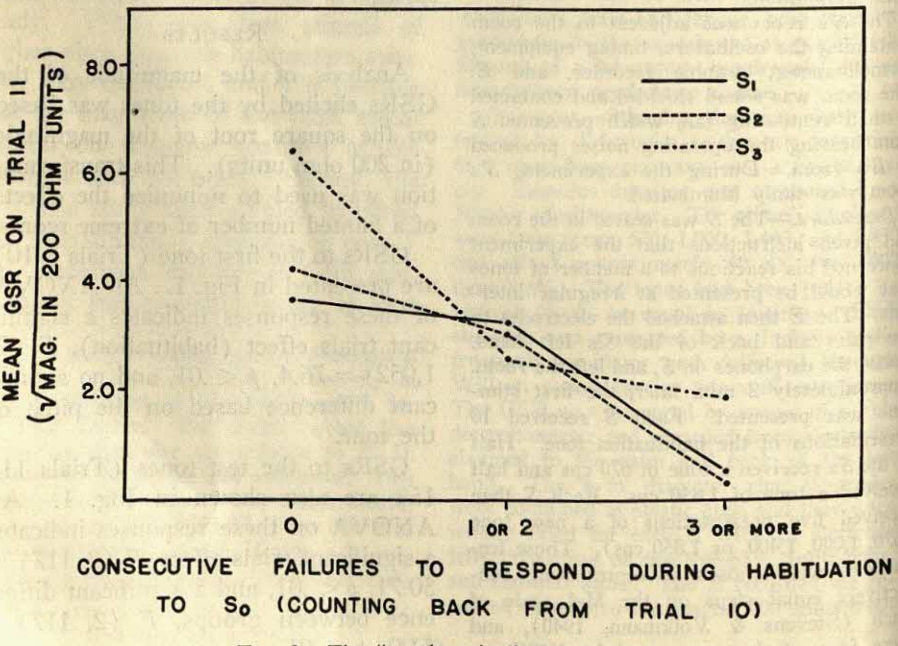


FIG. 3. The "overlearning" phenomenon.

ferences between the base levels of the groups.

Discussion

In regard to GSR magnitude to the first tone (S_0), Fig. 1 illustrates that there is a marked initial response to S_0. The determinants of the magnitude of this initial response are not known. However, it is of interest that the pitch of the tone (670 or 1,850 cps) was not a relevant factor. With repeated presentations of S_0, a significant decrement in the GSR magnitude occurred. Forty-six percent of the *S*s failed to give an observable response on Trial 10. When the new tone was presented (Trial 11) a GSR occurred, the magnitude of which was a function of the similarity in pitch of that new tone to the previously experienced tone (S_0) (Fig. 1 and 2). As the new tone was repeatedly presented (Trials 11–15), there was a decrement in response to the new tone.

The "overlearning" phenomenon (Fig. 3) indicates that continued stimulus presentations, even after *S* had ceased to give an observable response, increase the generalized habituation. Such an effect would not be predicted from a noninhibition theory, such as those based on the "expectancy" of *S*. According to expectancy theory, *S* ceases to respond to the first stimulus because he comes to expect its occurrence, and his response when the stimulus is changed is due to "surprise." If the failure of *S* to respond to S_0 is an index of expectancy, the more "nonresponse" trials, the greater should be his expectancy and the greater his surprise when a different stimulus occurs. Just the opposite effect occurs with regard to GSR magnitude. The *S*s were questioned at the end of the experimental sessions, and on no occasion was an *S* unaware that the stimulus had changed during the experiment.

Maltzman (e.g., Maltzman & Raskin, 1965) has observed consistent individual differences in the magnitude of the GSR to stimulus changes. In the present study, there was a significant correlation between *S*s' initial response to S_0 (Trial 1) and to the test stimulus (Trial 11) only when the greatest change in pitch occurred. This suggests that when inhibition is accrued, the amount of inhibition is not necessarily proportional to the initial magnitude of the response. The absence of a significant correlation between Trials 1 and 11 in two of the three groups gives added support to the validity of the "overlearning" effect, since it also occurred in these groups.

In regard to base-level resistances, during the presentation of S_0 the only significant effect was a gradual increase in resistance over trials. This effect is usually observed in GSR conditioning studies. During presentations of the test stimuli, there was again a significant increase over trials. In addition, there was a significant Groups × Trials interaction, probably due to the initial decrease in resistance for the $S_0 \rightarrow S_3$ group. This was the group that gave the largest response on Trial 11. All groups showed a similar drop in base-level resistance on Trial 2 following the large response on Trial 1. It is possible that this temporary decline is of psychological significance in that it reflects an alerting of the *S* or it may be due to incomplete recovery from the preceding response. There was no significant difference in base levels between groups, therefore differences in response magnitude during testing could not be attributed to base-level differences.

When a novel stimulus is presented, a GSR is elicited. This response has been labeled an orienting response (OR) by many investigators. This experiment illustrates that as a stimulus is repeatedly presented, an inhibitory phenomenon occurs which is manifested in a decrement of the elicited GSR (GSR-OR). This inhibitory process generalizes to a decreasing degree to other stimuli subsequently presented along a dimension of similarity.

The repeated presentation of a stimulus and a subsequent shift to one or more similar stimuli is inherent in the procedure for investigating stimulus generalization of association. The failure of many studies of stimulus generalization to find a gradient on the first few test trials (or cycles) which employ the GSR as the response measure (Hovland, 1937; Littman, 1949; Wickens, Schroder, & Snide, 1954) could be attributable to an interaction between the gradient of association and the gradient of partial OR inhibition. The rapid completion of the OR inhibition to the test stimulus would explain the rapidly emerging gradient of association.

REFERENCES

HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tones. *J. gen. Psychol.,* 1937, **17,** 279–291.

LITTMAN, R. A. Conditioned generalization of the galvanic skin response to tones. *J. exp. Psychol.,* 1949, **39,** 868–882.

MALTZMAN, I., & RASKIN, D. C. Effects of individual differences in the orienting reflex on conditioning and complex processes. *J. exp. res. Personal,* 1965, **1,** 1–16.

STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: A revised scale. *Amer. J. Psychol.,* 1940, **53,** 329–353.

WICKENS, D. D., & SCHRODER, H. M., & SNIDE, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.,* 1954, **47,** 52–56.

WILLIAMS, J. A. Novelty, GSR, and stimulus generalization. *Canad. J. Psychol.,* 1963, **17,** 52–61.

(Received June 25, 1966)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

The titles and authors of accepted papers are listed here. In 1965 this program was being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology and the outcome of this trial is presently being evaluated.

This listing represents the entire backlog of manuscripts accepted by this journal. They appear here 2 or 3 months after acceptance due to the printing lag. Such listing is intended to allow readers to become aware of research many months in advance of journal publication.

Efference and the Conscious Experience of Perception: Leon Festinger,* Clarke A. Burnham, Hiroshi Ono, and Donald Bamber: Department of Psychology, Stanford University, Stanford, California 94305.

Effects of Differential Instructions, Differential Feedback, and UCS Intensity on the Conditioned Eyelid Response: Harold D. Fishbein*: Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

Effects of Differential Instructions and Number of Acquisition Trails on Extinction and Reacquisition of the Conditioned Eyelid Response: Harold D. Fishbein*: Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

Effect of Sleep on Memory: Bruce R. Ekstrand*: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

Human Eyelid Conditioning as a Function of Interstimulus Interval: Harold D. Fishbein* and Mary LeBlanc: Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

Averaging Model Analysis of Set-Size Effect in Impression Formation: Norman H. Anderson*: Department of Psychology, University of California, San Diego, P. O. Box 109, LaJolla, California 92037.

(Continued on page 253)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 241–248

EFFECT OF STIMULUS DISTANCE AND AGE ON SHAPE CONSTANCY[1]

KATHLEEN A. MENEGHINI[2] AND H. W. LEIBOWITZ

Pennsylvania State University

Development of shape constancy as a function of age was investigated with distance as a parameter. In Experiment I, 60 adult Ss matched the shape of an inclined, circular test object with 1 of a series of comparison ellipses under either binocular or monocular viewing conditions, with the test object at 3, 15, or 30 ft. With binocular vision, the matches at 3 ft. represented a compromise between the true shape of the test object and its projected shape, while matches under monocular viewing correspond more closely to the retinal shape. With increasing distance, all matches approached the retinal shape. In Experiment II, 160 children (4–16 yr.) and an adult group (mean age = 19.4) made shape matches with the test object at either 3 or 15 ft. At 3 ft., shape constancy was inversely related to age, while at 15 ft. constancy was minimal at all ages. It was concluded that either the observed function is a result of variation in the interpretation of the instructions with increasing intellectual development or the adaptive value of shape constancy is diminished in importance as the maturing organism learns and utilizes additional cues for object identification.

The tendency to perceive the true shape of objects in spite of variation in their projected retinal image patterns resulting from varying angles of observation is known as shape "constancy" (Woodworth & Schlosberg, 1954). The biological significance of shape constancy is to provide stability of the perceptual world by emphasizing the permanent qualities of objects under various conditions of observation. Although it is well established that the tendency toward shape constancy is typical of adults (Thouless, 1931), there are relatively fewer data available for children. The principal development study was carried out by Klimpfinger who, while working in Brunswik's Vienna laboratory, reported that the tendency toward shape constancy increased asymptotically between the ages of 3 and 14 yr. (Klimpfinger, 1933a). Her results were interpreted as reflecting the influence of learning on shape perception in particular and the constancies in general (Postman & Tolman, 1949).

Later studies suggest that Klimpfinger's methodology may have inadvertently influenced her results. She used an elliptical test object with an axis ratio (ratio of vertical to horizontal axis) of 1.5. This particular shape has recently been shown to produce an anomalously greater tendency toward shape constancy than either a circular test object or an elliptical test object of even greater eccentricity (Leibowitz & Meneghini, 1966). Another possible complication results from the fact that Klimpfinger's matches were carried out with the test object at a distance of 16.4 ft. A number of studies published subsequently suggest that perceptual development proceeds

---

[1] Supported by Award 1-F1-MH-25, 495-01 and Grant MH-08061 from the National Institute of Mental Health of the National Institutes of Health.

[2] Now at University of Wisconsin, Madison, Wisconsin.

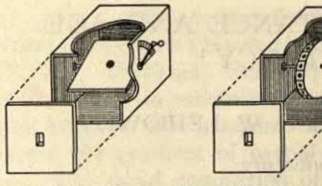

TEST OBJECT PRESENTATION
COMPARISON PRESENTATION

FIG. 1. Schematic diagram of the apparatus. (Left: Test-object presentation. The turntable on which the circular test object is mounted is adjustable with respect to tilt. Right: Comparison presentation. The comparison test objects range in axis ratio from .114 to 1.302 and may be viewed, one at a time, through the opening in the front of the box.)

from near observation distances, e.g., arm's length, and then only gradually extends to greater distances of the order of magnitude used by Klimpfinger (Vernon, 1962, Zeigler & Leibowitz, 1957). In view of these results, it is possible that Klimpfinger's data may simply reflect a manifestation of the development of perception at a distance and not be generally indicative of the developmental function of shape constancy.

The purpose of the present study is to investigate the development of shape constancy for a circular test object as a function of age with particular attention to the parameter of distance. Experiment I was designed to determine the effect of observation distance in order to ascertain whether, in fact, distance is a variable in shape matching. Adult *S*s were tested at distances of 3, 15, and 30 ft. Experiment II investigated the effect of age on shape constancy at different distances. Children ranging in age from 4 to 16 yr., as well as adults, were tested at observation distances of 3 and 15 ft.

EXPERIMENT I

Subjects.—The *S* population consisted of 60 male and female undergraduate students in introductory psychology courses who volunteered for this experiment in order to fulfill a course requirement.

Apparatus.—In order to test different *S* populations used in this study and the following study, it was necessary that the apparatus be portable. Therefore, both the stimulus object and comparison objects were mounted in boxes (2 ft. × 2 ft. × 2 ft.) painted flat black (reflectance 1.78%), and mounted on wooden bases which could be adjusted in height between 3–4 ft. (Fig. 1). This procedure insured that the environment immediately surrounding the stimuli was uniform at the various test locations.

A turntable was mounted in the center of the box which was used for presentation of the test object. The angle of orientation of the turntable with respect to *S*'s line of vision could be adjusted from 90° (vertical) to 10° (from the horizontal) by a handle on the side of the box. The turntable was covered with black flock paper (reflectance .85%) to insure a homogeneous surface. (For a description of this material see Nidetz, 1951). The *S* viewed the test object through a 2 × 3 in. opening cut from the front face of the box and located 12 in. from the test object. The geometry of the construction was such that *S* never saw the edges of the turntable. The test object was a .12 in. thick white plastic circle (reflectance 62%) with a diameter 1.2 in. Its edges were painted black. In effect, *S* viewed a cylinder on a slanted background under an illuminance, provided by fluorescent lamps, of 38 ftc. The illuminance in the room could not be precisely controlled since testing was carried out at several locations. This varied between approximately 10 and 100 ftc.

The comparison objects consisted of 19 ellipses with axis ratios ranging from .114 to 1.302 in approximately equal steps. Each of these was mounted on one face of a wooden disk (diameter = 23 in.) cut in the shape of a polygon and covered with black flock paper. Rotation of the disk permitted viewing of each of the comparison objects independently through the 2 in. × 3 in. opening in the front of the box. The comparison objects were also made of .12 in. thick white plastic with a horizontal axis of 1.2 in. The wheel was driven by an electric motor and could be controlled by *S* with the aid of buttons which rotated the wheel and determined the direction of rotation.

Design.—The *S*s were tested either monocularly or binocularly with the test object at distances of 3, 15, or 30 ft. For the binocular condition, both eyes were used at all times.

TABLE 1

Mean Matched Axis Ratios as a Function of Angle of Inclination for Binocular and Monocular *O*s at Three Distances

Experimental Cond.	Angle of Inclination			
	10°	28°	45°	90°
Binoc. 3 ft.				
$\bar{X}$	.481	.758	.914	.996
SD	.235	.090	.085	.052
Binoc. 15 ft.				
$\bar{X}$	.204	.552	.749	1.025
SD	.066	.037	.050	.052
Binoc. 30 ft.				
$\bar{X}$	.169	.538	.769	.983
SD	.070	.110	.089	.105
Monoc. 3 ft.				
$\bar{X}$	.314	.643	.792	1.015
SD	.182	.116	.145	.079
Monoc. 15 ft.				
$\bar{X}$	.253	.607	.742	.985
SD	.100	.128	.094	.148
Monoc. 30 ft.				
$\bar{X}$	.244	.560	.726	1.027
SD	.127	.161	.060	.103

In monocular testing (preferred eye) the test and comparison objects were both viewed monocularly. Ten *S*s served in each of the six experimental conditions and viewed the test objects at angles of orientation of the turntable of 10°, 28°, 45°, and 90° (vertical). The order of the experimental conditions was determined by use of a table of random numbers and *S*s were assigned to the conditions in the order in which they arrived for testing.

Procedure.—The *S*s were seated with the comparison box at a distance of 3 ft. to their left and the test object at a distance of either 3, 15, 30 ft. to their front. Both boxes were adjusted so that the horizontal axes of the test and comparison objects were at eye level.

Instructions.—"Would you please look at the white object in front of you and then look at the the white object on your left. Are they the same size?[3] Now would you look to your left, please. On your left, there are many different shaped objects. Your task will be to look at the shape of the object presented in front of you and then to find a matching shape for it from one of these on your left."

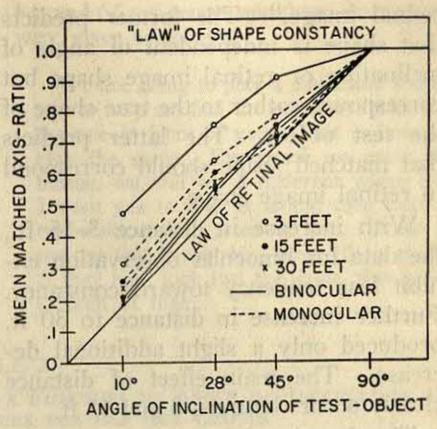


Fig. 2. Mean matched axis ratio as a function of angle of inclination of a circular test object. (Data were obtained at three distances of observation for monocular and binocular observation with adult *O*s.)

The *S*s were then given instructions as to the control of the apparatus, were shown a number of the comparison ellipses, and told to take as much time as was necessary to make satisfactory shape matches. The testing was usually completed within 5 min.

Results.—The data, the mean matched axis ratio as a function of angle of inclination, are presented in Table 1 and plotted in Fig. 2 for the three distances of observation. The curve for "binocular 3 ft." is typical of data obtained from adult *O*s in shape-matching studies, i.e., the matches represent a compromise between theoretical predictions based on the "law of shape constancy" and the "law of the retinal image." The former predicts that shape is independent of angle of inclination or retinal image shape but corresponds rather to the true shape of the test object. The latter predicts that matched shape should correspond to retinal image shape.

---

[3] In testing at distances of 15 and 30 ft., there exists the possibility that for any child or individual adult *S*, size constancy may be incomplete (Zeigler & Leibowitz, 1957). In order to eliminate size as a variable, *S* was first shown the test object at an angle of 90° and asked to compare its size with the circular comparison object. If *S* reported that the test object appeared either smaller or larger than the comparison object, the test object was then replaced with either a larger or smaller test object until *S* reported that the test object and comparison object appeared to be the same size. In effect, this procedure resulted in all shape comparisons being made for disks of subjectively equal size.

With increase in distance 3–15 ft., the data for binocular observation exhibit less tendency toward constancy. Further increase in distance to 30 ft. produced only a slight additional decrease. The main effect of distance clearly occurs between 3 and 15 ft.

The function labeled "monocular 3 ft." is typical of studies in which monocular observation has been employed. The data exhibit a somewhat lower tendency toward constancy that those obtained under binocular conditions (Leibowitz, Bussey, & McGuire, 1957). With increase in distance, the tendency toward constancy is slightly reduced.

A three-way analysis of variance indicated that the differences between the matches obtained under binocular and monocular conditions were not significantly different $F\ (1, 54) = .96$. For both binocular and monocular conditions, an increase in distance results in a significant decrease in the tendency toward constancy, $F\ (2, 45) = 13.13$, $p < .001$.

Discussion.—The present results clearly indicate that the tendency toward shape constancy is inversely related to distance for both binocular and monocular observation conditions. This implies that the effect of distance is functionally equivalent to elimination of the cues which ordinarily subserve shape constancy. Whether these cues are provided by slant, texture, etc., is of considerable theoretical importance and remains a matter for further experimental analysis. However, the data confirm the assumption that distance is a variable in shape matching. The possibility therefore exists that the previous results of Klimpfinger may have been unduly influenced by her use of a relatively extended distance of observation.

EXPERIMENT II

The principal objective of Exp. II was to determine the effect of age on shape matching. Since distance has been shown to be a significant factor, matches were carried out at observation distances of both 3 and 15 ft. to determine (*a*) if the tendency toward shape constancy varies with age and (*b*) whether the various age groups are differentially affected by observation distance.

Subjects.—The *S* population included nine age groups consisting of 10 males and 10 females each. The mean age in years and *SD* of the groups were as follows: 4.54 (.44); 5.59 (.99); 6.54 (.25); 8.54 (.37); 10.52 (.52); 12.56 (.85); 14.72 (.33); 16.58 (.19); and 19.37 (.76).

The youngest age group was obtained from the University Nursery School and the oldest from elementary psychology classes. The intermediate age groups were students from local elementary, junior, and senior high schools.[4] The method of selection was to identify those children closest to the desired age (plus 6 mo.) who were available during the scheduled testing periods. The adults were volunteers who participated in order to satisfy a course requirement.

Four *S*s were replaced in the 4.5 age group, two because they would not remain to complete the experiment, and two because they did not understand the concept of circle. Four *S*s were replaced in the 5.6 age group because they could not understand the instructions. One was replaced in the 6.5 age group due to inattention, and one in the 8.5 age group because of failure to understand the instructions.

Design and procedure.—All *S*s were tested binocularly with the test object placed at

---

[4] The authors wish to express their appreciation to Winona Morgan, Director of the University Nursery School, Bruce Brummitt, Superintendent of the State College Area Schools, the State College Area School Board, and the principals and staff of the State College Area Schools for their cooperation in making these *S*s available.

either 3 or 15 ft. Each *S* was tested at four angles of inclination of the test object: 10°, 28°, 45°, and 90° (vertical). The order of presentation of the angles was determined by a table of random numbers, and the order of the testing distance was alternated for successive *S*s. The apparatus was the same as in Exp. I.

Instructions.—The instructions for *S*s ranging in age from 12.5 and older were identical to those used in Exp. I. For the younger *S*s, comparable instructions with simpler phrasing were used. For the 4.5, 5.6, and 6.5 yr. groups additional instructions were given as follows:

> We are going to play a game and I want you to help me find the one that looks like the one in front of you. This is how we play the game. Everytime we press the button, we will see a different white one. I want you to look at it and then look at the white one in front of you. Point to the one that is rounder—more like a circle. If they both look just the same, then point to both of them.

TABLE 2

MEAN MATCHED RESPONSE AXIS RATIOS AS A FUNCTION OF ANGLE OF INCLINATION AT TWO OBSERVATION DISTANCES FOR ALL AGE GROUPS

Age	Distance	Angle of Inclination			
		10°	28°	45°	90°
4.5	3 ft.				
	$\bar{X}$	.948	.882	.895	1.020
	SD	.070	.097	.099	.062
	15 ft.				
	$\bar{X}$	.284	.559	.767	.964
	SD	.263	.155	.100	.066
5.6	3 ft.				
	$\bar{X}$	.952	.946	.998	1.007
	SD	.059	.124	.093	.048
	15 ft.				
	$\bar{X}$	.335	.509	.732	.959
	SD	.250	.110	.108	.070
6.5	3 ft.				
	$\bar{X}$	.706	.866	.956	1.003
	SD	.315	.104	.068	.011
	15 ft.				
	$\bar{X}$	.253	.546	.704	.950
	SD	.179	.227	.072	.072
8.5	3 ft.				
	$\bar{X}$	.629	.816	.901	.966
	SD	.187	.096	.080	.061
	15 ft.				
	$\bar{X}$	.280	.514	.773	.957
	SD	.224	.139	.121	.060
10.5	3 ft.				
	$\bar{X}$	.592	.732	.888	.994
	SD	.238	.167	.122	.039
	15 ft.				
	$\bar{X}$	.275	.523	.706	.922
	SD	.075	.086	.131	.107
12.6	3 ft.				
	$\bar{X}$	.481	.769	.856	.947
	SD	.202	.082	.125	.079
	15 ft.				
	$\bar{X}$	.192	.477	.682	.972
	SD	.067	.090	.104	.054
14.7	3 ft.				
	$\bar{X}$	.546	.785	.897	.973
	SD	.361	.206	.155	.067
	15 ft.				
	$\bar{X}$	.240	.512	.706	1.020
	SD	.186	.066	.068	.033
16.6	3 ft.				
	$\bar{X}$	.411	.738	.857	.981
	SD	.143	.133	.098	.051
	15 ft.				
	$\bar{X}$	.171	.512	.690	.997
	SD	.105	.066	.099	.064
19.4	3 ft.				
	$\bar{X}$	.426	.725	.903	1.014
	SD	.173	.154	.117	.029
	15 ft.				
	$\bar{X}$	.269	.549	.715	.957
	SD	.080	.090	.073	.060

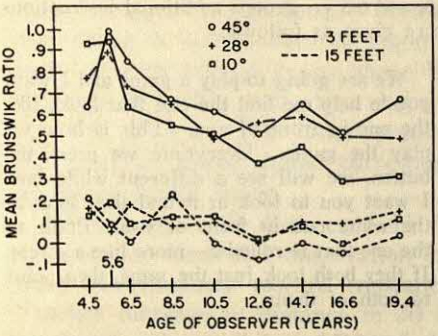


FIG. 3. Mean Brunswik ratio (percent constancy) as a function of the age of *O* for three angles of inclination at observation distances of 3 and 15 ft.

Most of the children in the 4.5, 5.6, and 6.5 yr. groups reported that two or more of the comparison objects were identical in shape to the test object. In this case, the mean axis ratio of all indicated ellipses were recorded as the child's match. Children of this age required 15–30 min. to complete the experiment, while the older age groups required about 5 min.

Results.—The mean matched Brunswik ratio as a function of age for inclinations of 10°, 28°, and 45° at observation distances of 3 and 15 ft. are presented in Table 2 and plotted in Fig. 3. The data for 90°, which is essentially a control condition, have not been plotted. The Brunswik ratio is used as a quantitative expression of the percentage of constancy in order to facilitate comparison of the tendency toward constancy independent of angle of inclination.[5]

At the 3-ft. observation distance, the youngest children manifest nearly complete constancy for all angles of inclination. However, with increase in age the tendency toward constancy decreases such that the matches of the adults represent a compromise between the theoretical predictions based on the law of shape constancy and the law of the retinal image. There is a slight variation in this function related to angle of observation. The maximum for the 10° angle is observed at 4.5 yr. of age, while the maximum for the 28° and 45° inclinations occur at the 5.6 age level. After these ages, all functions decrease systematically, rapidly at first and then more slowly. For all angles of observation, a slight rise is noted at age 14.7. A three-way analysis of variance revealed that the variables of age, F (8, 162) = 5.34; distance, F (1, 162) = 243.26, and angle of inclination, F (3, 486) = 842.54, are all significant at beyond the .001 level of confidence.

For the 15 ft. observation distance, all data lie close to the law of the retinal image with only a slight decrease in the tendency toward constancy with increasing age.

Discussion.—The implications of the present data may be treated independently with respect to the variables of distance and age. Regarding the effect of distance for the adults, it is clear that shape constancy is markedly reduced as observation distance is increased and is virtually nonexistent at 30 ft. As suggested by the data on monocular shape matching, one factor would seem to be the reduction in the magnitude of retinal disparity. However, even with monocular vision, increasing distance reduces the tendency toward shape constancy. Since a number of cues in addition to disparity are reduced as distance is increased, e.g., texture, cues for slant, etc., it is not possible to specify which monocular cues or combination of cues are responsible for the observed effects. These results would provide an interesting and theoretically important point of departure for further analysis.

The more interesting aspect of the present study is the observed decrease in the tendency toward shape constancy as a function of age at near distances, and the marked reduction in shape constancy for all age groups at the far distances. It should be pointed out that although the 15-ft. testing distance is an experimental condition, it is also, in effect, a control condition. If the instructions were misunderstood by the younger age groups, it would be expected that their matches would be similar at both distances. These results, showing that shape constancy decreases with age at the 3-ft. observation distance and is minimal at a distance of 15 ft. for *S*s of all ages, are particularly striking in view of Klimpfinger's (1933a) data which led to the generalization that shape constancy increases with age. They are also of general interest in relation to other aspects of perceptual constancy, since there is evidence that size and brightness constancy both increase with age (Postman & Tolman, 1949; Zeigler & Leibowitz, 1957). The observed effect of distance justifies the concern with Klimpfinger's choice of a 16.4-ft. test-object distance, and points to the importance of controlling this variable in studies of perceived shape.

---

[5] Calculated as (Axis ratio of *S*s match) —(axis ratio of retinal image)/(True axis ratio of test object, i.e., 1.00)—(axis ratio of retinal image).

The basis for the discrepancy between Klimpfinger's results and the present data is not clear. Her use of a relatively large elliptically shaped test object (axes 20 × 30 cm.) as compared to the smaller round (diameter 3 cm.) stimulus in the present study precludes direct comparison of the data. Additional information regarding the effect of test-object size and shape, distance, and the interactions among these variables as a function of age would be necessary before the observed differences can be satisfactorily understood.

The present data indicating a decrease in shape constancy with age are in agreement with a recent study by Vurpillot (1964) who reports a similar trend 7–12 yr. of age. In any consideration of the basis for the difference between the data of Klimpfinger and that of Vurpillot and the present study, as well as the theoretical significance of the functions obtained, the variable of instructions must be taken into account. A number of studies have clearly indicated that shape matching is significantly influenced by instructions (Eissler, 1933; Joynson & Newson, 1962; Klimpfinger, 1933b; Landauer, 1964). When *S* is requested to adopt an "analytic" attitude, the matches tend towards a retinal image prediction. Instructions to match in terms of "objective" shape produce matches more nearly in accord with the law of shape constancy. Although instructions were not specifically varied by Klimpfinger (1933a) or in the present study, the possibility cannot be overlooked that the same objective instructions are interpreted differently at various age levels. In a study of the effect of intelligence level on matched shape using the same objective instructions, the tendency toward shape constancy was found to decrease as a function of intelligence level (Leibowitz, Waskow, Loeffler, & Glaser, 1959). In view of the similarity of the trends obtained as a function of age and intelligence, there is a possibility that both functions reflect the adoption of an increasingly analytic attitude on the part of the older and/or more intelligent *S*s. Unfortunately, it is not possible to investigate perceived shape directly. Experimental studies must deal with matched shape which may be influenced, in a yet not clearly understood manner, by *S*'s reactions to the instructions.

Assuming that the present data are not merely a result of an interaction between age and instructions and that they do, in fact, represent the functional relationship between shape constancy and age, it is interesting to consider the significance of this trend. The present data are not inconsistent with the point of view that shape perception plays a decreasingly important role in perception with increasing age or intellectual development. For a young child who is unable or is just beginning to conceptualize or verbalize about his environment, shape constancy would be adaptive in that his ability to recognize objects would be increased by the invariant relationship between the physical properties of objects and his perception of them. However, for the maturing organism, with the development of analytic abilities, conceptualization, verbal behavior, and the awareness of a rich base of cues, reliance upon the early developed shape constancy is diminished, although not completely eliminated, by the establishment of alternative mechanisms. For an immature organism shape is probably one of the few cues available for object identification; for the mature organism, shape is only one of the many available cues. The assumption is that shape constancy represents a primitive adaptive technique which is strongly manifested in animals (Coburn, 1914; Locke, 1938; Zeigler & Leibowitz, 1957), adults of low intelligence (Leibowitz et al., 1959) and in the present case, normal young children. The adaptive value of high shape constancy is, however, diminished in importance as the maturing organism learns to utilize additional cues in his perceptual world.

Whether the function relating matched shape to age is determined by a varying reaction to the instructions or by a diminution in the relative importance of shape constancy cannot be resolved by the present study, the data of which are empirically consistent with both interpretations. The present results do emphasize the importance of variables such as distance and demonstrate that under at least one set of experimental conditions, the tendency toward constancy may decrease with age. Determination of the generality of this function is relevant not only for the understanding of shape matching behavior, but in relation to the general importance of the constancies in perception.

REFERENCES

Coburn, C. A. The behavior of the crow, *Coruus Americanus Aud., J. Anim. Behav.*, 1914, **4**, 185–201.

Eissler, K. Gestaltkonstanz der Sehdinge bei Variation der Objekte und Ihrer Einwirkungsweise auf den Wahrnehmenden. *Arch. ges. Psychol.*, 1933, **88**, 487–550.

Joynson, R. B., & Newson, L. J. The perception of shape as a function of inclination. *Brit. J. Psychol.*, 1962, **53**, 1–15.

Klimpfinger, S. Die Entwicklung der Gestaltkonstanz vom Kind zum Erwachsenen, *Arch. ges. Psychol.*, 1933, **88**, 599–628. (a)

Klimpfinger, S. Über den Einfluss von intentionaler Einstellung und Übung auf Gestaltkonstanz. *Arch. ges. Psychol.*, 1933, **88**, 551–598. (b)

Landauer, A. A. The effect of viewing conditions and instructions on shape judgments. *Brit. J. Psychol.*, 1964, **55**, 49–57.

Leibowitz, H. Bussey, R., & McGuire, P. Shape and size constancy in photographic reproductions, *J. Opt. Soc. Amer.*, 1957, **47**, 658–661.

Leibowitz, L., & Meneghini, K. Shape perception for round and elliptically shaped test-objects. *J. exp. Psychol.*, 1966, **72**, 244–249.

Leibowitz, H., Waskow, N., Loeffler, N., & Glaser, F. Intelligence level as a variable in the perception of shape. *Quart. J. exp. Psychol.*, 1959, **11**, 108–112.

Locke, N. M. Perception and intelligence: Their phylogenetic relation. *Psychol. Rev.*, 1938, **45**, 335–345.

Nidetz, M. A light absorbent surface. *Amer. J. Psychol.*, 1951, **64**, 109–110.

Postman, L., & Tolman, E. C. "Brunswik's probabilistic functionalism," In S. Koch (Ed.), *Psychology: A study of a science*, I. New York: McGraw-Hill, 1949. Pp. 539–541.

Thouless, R. H. Phenomenal regression to the "real" object. I. II. *Brit. J. Psychol.*, 1931, **21**, 339–359; **22**, 1–30.

Vernon, M. D. *The psychology of perception.* Baltimore: Penguin, 1962.

Vurpillot, E. Perception et représentation dans la constance de la forme. *L'Année Psychol.*, 1964, No. 1, 61–82.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Henry Holt, 1954.

Zeigler, H. P., & Leibowitz, H. Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.*, 1957, **70**, 106–109.

Zeigler, H. P., & Leibowitz, H. A methodological study of shape constancy in the Rhesus monkey. *J. comp. physiol. Psychol.*, 1958, **51**, 155–160.

(Received June 27, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 249–253

EFFECT OF THE ANGLE OF TILT OF THE INSPECTION FIGURE ON THE MAGNITUDE OF A KINESTHETIC AFTEREFFECT

RAY OVER

University of Otago

The magnitude of a kinesthetic aftereffect, specified in terms of the difference between post- and preinspection settings of a bar to the apparent horizontal, has been measured as a function of the angle of tilt of the bar during the inspection period. Postinspection settings were found to be displaced from preinspection settings in the direction of the tilt of the inspection figure at all angles between 15° and 75°. The maximum aftereffect was found at 60°. These results differ from those found by other Es for visual aftereffects and indicate that aftereffects involving judgments of tilt cannot be considered solely in terms of the dimensional relationship between the inspection figure and the test figure. It is suggested that an explanation of an aftereffect must be sought in terms of the operating characteristics of the sensory system involved in inspection and making spatial judgments.

A physically vertical line looks to be tilted slightly following inspection of a tilted line. The direction and magnitude of the aftereffect induced by inspection are a function of the angle of tilt of the inspected contour (Gibson & Radner, 1937; Morant & Harris, 1965). Clockwise tilts of less than 45° from the vertical lead to the vertical subsequently appearing to be tilted slightly in an anticlockwise direction, while clockwise tilts of between 45° and 90° produce clockwise displacements. Gibson and Radner (1937) have called the former aftereffects, which are maximal when the inspection figure is tilted about 15°, the "direct effect," and the latter, which are maximal at about 75°, the "indirect effect." Visual aftereffects are not obtained when the inspection figure is at a tilt of 0°, 45°, or 90° from the vertical.

Gibson (1937b) has suggested that the horizontal and the vertical are norms of visual space and that during prolonged inspection of a tilted line there is adaptation toward the adjacent spatial norm. During adaptation there is a displacement of the psychophysical correspondence between physical tilt and experienced tilt, but the coordering of the stimulus dimension and the experience continuum is maintained (Gibson, 1937b, 1959). A line tilted to the right of the vertical will appear more vertical during inspection (i.e., it will shift leftwards), and a line subsequently presented at the vertical will look tilted to the left. In these terms the direction of an aftereffect will be dependent on both the norm toward which adaptation has occurred and the orientation of the postinspection figure relative to this norm; the magnitude of an aftereffect will be determined by the amount of adaptation that has taken place. A line tilted 45° would lie midway between two norms and thus would not adapt during inspection; no aftereffect is found following inspection of a line at this tilt (Gibson & Radner, 1937). Gibson (1937b) has suggested that the above analysis can be applied to kinesthetic aftereffects produced by inspection of a tilted figure. This issue is examined in the present experiment.

Gibson and Backlund (1963) have studied a kinesthetic aftereffect of the vergence of two surfaces; *S*'s hands were placed for a period of time on two palmboards set at a given angle of vergence and *S* was subsequently required to adjust the boards so that they felt parallel. The effects of inspection angles ranging 11.25°–90° were examined; when the angle was 90° each hand was tilted 45° from the vertical. The aftereffect found at each angle of tilt was of the "direct effect" type and aftereffects of greatest magnitude were obtained when the angle of vergence was 90°. This latter finding suggests that visual and kinesthetic aftereffects are not similarly influenced by the tilt of the inspection figure. Gibson (1937a) has claimed, however, that the vergence figure is not simply a compound form of the figure used by Gibson and Radner (1937) to study visual aftereffects. In addition Morant and Harris (1965) have found that "direct effect" displacements are obtained at all inspection angles between 15° and 75° when *S* is required to set a previously inspected line so that it looks parallel to an objectively vertical line.

Gibson (1937b) has described a measure of kinesthetic aftereffect in which a blindfolded *S* is required to move his hand along a tilted bar for a period of time before setting the bar so that it feels horizontal. In both this task and the situation in which visual aftereffects were measured by Gibson and Radner (1937) there is an active inspection process (hand movements in one case and eye movements in the other) and *S* is required to make an absolute, rather than a relative, judgment following inspection. Gibson (1937b, 1959) has implied that visual and kinesthetic aftereffects have similar sets of determinants. The present experiment examines kinesthetic aftereffects obtained as the angle of the tilt of the inspection figure is varied from 0° to 90°.[1] It is asked whether both "direct effects" and "indirect effects" are found following kinesthetic inspection.

METHOD

Apparatus.—A wooden bar 24-in. long and 1-in. thick was mounted at eye height 14-in. in front of *S*. The *S* was blindfolded and his head was supported by a chin rest during testing. When instructed he was required to move his right hand along a center 12-in. section of the bar, the limits of which were marked by tape, in time with a metronome (84 beats/min.). The bar could be set at any angle of tilt by *E*. The *S* was able to change the tilt of the bar by operating rotary controls with his left hand. Settings of the bar could be read by *E* to the nearest .25° from a protractor and marker mounted at the rear of the frame supporting the bar.

Subjects.—There were seven groups, each of eight *S*s. The *S*s were students in an introductory course in psychology.

Procedure.—Each group was tested with the bar being at only one angle of tilt during inspection. The seven angles of tilt, relative to the physical horizontal, which were used were: 0°, 15°, 30°, 45°, 60°, 75°, and 90°. For half of the *S*s in each group the bar was tilted in the clockwise direction from the horizontal and for half in the anticlockwise direction. For the former *S*s inspection of the bar involved an inward rotation of the arm, and for the latter *S*s an outward rotation of the arm.

The *S*s were tested individually. Each *S* made four postinspection settings to the apparent horizontal by operating rotary controls with the left hand while moving the right hand across the bar in time with the metronome. Each postinspection setting was preceded by an inspection period of 60 sec. during which *S* moved his right hand across the bar in time with the metronome. Four preinspection settings to the apparent horizontal preceded each inspection period. The settings were made from one of two starting positions (15° from the horizontal in the clockwise direction, and 15° in the anticlockwise direction). A postinspection setting and the preceding four preinspection settings were all obtained from the same starting position. Each *S* made two postinspection settings from each starting position. The order of testing was varied across *S*s. The *S*s were required to complete all settings between 5 sec. and 10 sec. of being instructed. Following each postinspection setting *S* was given a rest of 6 min. before the first preinspection setting of the next sequence began.

---

[1] Gibson and Radner (1937) have examined the effect of angle of tilt on aftereffects measured by obtaining judgments of verticality. Morant and Harris (1965) have shown that direct and indirect effects are also obtained when *S* is required to make visual judgments of the apparent horizontal.

RESULTS

Aftereffects were measured in terms of the difference between a postinspection setting and the mean of the preceding four preinspection settings. If the postinspection setting was displaced from the mean preinspection measure in the direction of the tilt of the inspection figure, the aftereffect was scored +; if in the opposite direction, − was scored. Differences in a clockwise direction were scored + and in an anticlockwise direction − for the group for which the bar was physically horizontal during inspection. Table 1 shows the mean aftereffects obtained with each angle of tilt of the inspection figure; the aftereffects found when inspection involved an inward rotation of the arm are shown separately from those found when inspection required an outward rotation of the arm.

Aftereffects obtained from the two starting positions did not differ significantly in magnitude, t (111) = .16. Postinspection settings differing significantly ($p < .01$) from preinspection settings in the direction of the tilt of the inspection figure were obtained at all angles of tilt between 15° and 75°. Aftereffects were in the + direction for all *S*s in the above groups. The differences between the four measures of aftereffect obtained from each *S* were not significant, F (3, 156) = 2.14. By an analysis of variance,[2] the effect of angle of tilt of the inspection figure, F (5, 36) = 5.62, $p < .01$, and the direction of rotation of the arm when in the position required for inspection, F (1, 36) = 13.15, $p < .01$, were found to be significant. The interaction, Angle of Tilt × Direction of Rotation, was not significant, F (5, 36) = .36.

DISCUSSION

From the account of aftereffects given by Gibson (1937b, 1959), it might be

TABLE 1

MEANS AND *SD*s OF DIFFERENCES BETWEEN POST- AND PREINSPECTION SETTINGS

Inspection Cond.	Angle of Tilt of Bar during Inspection						
	0°	15°	30°	45°	60°	75°	90°
Inward $\bar{X}$	+0.80	+4.52	+6.22	+8.06	+10.01	+7.81	+4.31
Rotation *SD*	1.72	1.51	1.26	1.01	1.39	2.30	6.68
Outward $\bar{X}$	+0.80	+2.75	+4.96	+5.45	+ 6.96	+3.81	+0.70
Rotation *SD*	1.72	2.05	1.21	0.43	4.08	1.93	1.71

---

[2] Measures obtained when the inspection figure was tilted 0° have not been included in this analysis of variance; direction of rotation cannot be specified as inward or outward at this tilt. Measures obtained when the bar was vertical (90°) have been included in that half of the *S*s inspected with the palm of the hand inward relative to the body and half with the palm outward.

expected that the direction and magnitude of kinesthetic and visual aftereffects would be similarly influenced by the angle of tilt of the inspection figure. "Direct effects" and "indirect effects" should be found for each modality and the maximums should be similarly located. The results of the present experiment show that only "direct effects" are found following kinesthetic inspection of a tilted figure; i.e., when postinspection settings to the horizontal differed from preinspection settings the difference was always in the direction of the tilt of the inspection figure. Whereas visual aftereffects of greatest magnitude are found following inspection of a line tilted only slightly from a spatial norm (Gibson & Radner, 1937; Morant & Harris, 1965), kinesthetic aftereffects of greatest magnitude were obtained when the inspection figure was tilted 60° from the horizontal. In addition the kinesthetic aftereffects obtained at almost every angle of tilt of the inspection figure were of much greater magnitude than the largest visual aftereffects which have been reported.

The fact that aftereffects in the two modalities are differently controlled by the angle of tilt of the inspection figure suggests than an explanation of an aftereffect is to be sought in terms of the properties of the sensory system involved in inspection and in the making of spatial judgments rather than simply in terms of the dimensional relationships between the inspection figure and the test figure. This claim is reinforced by the finding that, at each angle of tilt of the inspection figure, larger kinesthetic aftereffects are obtained when the arm is in an inward rotated position than in an outward rotated position during inspection.

Sutherland (1961) and Day (1962) have made use of data available on the organization of receptive fields in the visual system (Hubel & Wiesel, 1959) to offer an account of visual aftereffects involving judgments of tilt. They suggest that, when the tilted inspection figure has stimulated a field maximally responsive to a vertical line, the usual excitatory "on" discharge produced when a vertical line is introduced will be inhibited by the "off" activity produced by removing the inspection figure. The discharge produced by the vertical line (the postinspection figure) would be that normally signaled by a slightly tilted line.

An account of kinesthetic aftereffects involving judgments of tilt can be offered in terms of the organization of receptive fields in the kinesthetic system. Many of the thalamic cells related to a given joint have excitatory angles greater than half of the total angular movement possible for the joint. Some are maximally active at one limit of movement of the joint and others at the other limit (Mountcastle, Poggio, & Werner, 1963). Both classes of cells discharge at a low rate when the joint is in the midposition of its total angular range. A decrease in the activity of cells sensitive to displacement in one direction from the midposition relative to the activity of cells sensitive to displacement in the opposite direction signals a shift in the position of the joint. Restoring a joint to the midposition after it has been maintained away from the midposition for a period of time results in an inhibition of activity of the cells previously active (Mountcastle & Powell, 1959). If the extent of inhibition is a function of the previous rates of discharge of the cells, it can be seen that the shift in position signaled when the joint is restored to the midposition would be directly related to the prior angular displacement of the joint from the midposition up to the limits of the excitatory angle of cells related to the joint.

REFERENCES

Day, R. H. Excitatory and inhibitory processes as the basis of contour shift and negative after-effect. *Psychologia,* 1962, **5,** 185–193.

Gibson, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *J. exp. Psychol.,* 1937, **20,** 553–569. (a)

Gibson, J. J. Adaptation with negative after-effect. *Psychol. Rev.,* 1937, **44,** 222–244. (b)

Gibson, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A Study of a Science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

Gibson, J. J., & Backlund, F. A. An after-effect in haptic space perception. *Quart. J. exp. Psychol.,* 1963, **15,** 145–154.

Gibson, J. J., & Radner, M. Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies. *J. exp. Psychol.,* 1937, **20,** 453–467.

Hubel, D. H., & Wiesel, T. N. Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.,* 1959, **148,** 574–591.

Morant, R. B., & Harris, J. R. Two different after-effects of exposure to visual tilts. *Amer. J. Psychol.,* 1965, **78,** 218–226.

Mountcastle, V. B., Poggio, G. F., & Werner, G. The relationship of thalamic cell response to peripheral stimuli varied over an intensive continuum. *J. Neurophysiol.,* 1963, **26,** 807–834.

Mountcastle, V. B., & Powell, T. P. S. Central nervous mechanisms subserving position sense and kinesthesis. *Bull. Johns Hopkins Hosp.,* 1959, **105,** 173–200.

Sutherland, N. S. Figural after-effects and apparent size. *Quart. J. exp. Psychol.,* 1961, **13,** 222–228.

(Received June 27, 1966)

(Continued from page 240)

Effect of Tests Without Feedback and Presentation-Test Interval in Paired-Associate Learning: Thomas K, Landauer* and Lynn Eldridge: Laboratory of Human Development, Stanford University, Stanford, California 94305.

Spontaneous Recovery of List 1 Responses in the A-B, A-C Paradigm: Eli Saltz* and Helen Hamilton: Department of Psychology, Wayne State University, Detroit, Michigan 48202.

Amount of Uncertainty Associated with Decoding in Free Recall: Marvin R. Mueller,* Ed M. Edmonds, and Selby H. Evans: Department of Psychology, Augusta College, Augusta, Georgia 30904.

Time and Effort as Determiners of Time Production Error: Lawrence R. Boulter* and Mortimer H. Appley: Department of Psychology, York University, Toronto 12, Ontario, Canada.

Pacing of Rehearsal in Sequential Short-Term Memory: Richard A. Monty,* Robert Karsh, and Harvey H. A. Taub: Research Psychologist, Supporting Research Laboratory, U. S. Army Engineering Laboratories, Aberdeen Proving Ground, Maryland 21005.

Comparison of Anticipation and Recall Methods in Paired-Associate Learning: Charles N. Cofer,* Florence Diamond, Richard A. Olsen, Judith S. Stein, and Howard Walker: Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

Latency of Imaginal and Verbal Mediators as a Function of Stimulus and Response Concreteness-Imagery: John C. Yuille* and Allan Paivio: Department of Psychology, Middlesex College, University of Western Ontario, London, Ontario, Canada.

Auditory and Cutaneous Temporal Resolution of Successive Brief Stimuli: George A. Gescheider*: Department of Psychology, Hamilton College, Clinton, New York 13323.

Imagery and Association Value in Paired-Associate Learning: Allan Paivio* and Stephen A. Madigan: Department of Psychology, University of Western Ontario, Middlesex College, London, Canada.

Effects of Interdimensional Training on Stimulus Generalization: II. Within-Subjects Design: Joseph Lyons and David R. Thomas*: Department of Psychology, 215 Kent Hall, Kent State University, Kent, Ohio 44240.

Meaningfulness and Abstractness in Short-Term Memory: John G. Borkowski* and Howard C. Eisner: Department of Psychology, Oberlin College, Oberlin, Ohio 44074.

Retroactive Inhibition of the R-S Associations in the A-B, B-C and A-B, C-B Paradigms: Chiu C. Cheung and L. R. Goulet*: Department of Psychology, West Virginia University, Morgantown, West Virginia 26506.

Temporal Course of Perception in an Immediate-Recall Task: Doris Aaronson*: Center for Cognitive Studies, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, Massachusetts 02138.

(Continued on page 293)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 254–258

UNLEARNING OF SPECIFIC ASSOCIATIONS IN THE A-B, C-D PARADIGM[1]

JOHN P. HOUSTON
University of California, Los Angeles

The 1st of 2 experiments asked whether specific as well as contextual 1st-list associations are unlearned in the A-B, C-D paradigm and whether spontaneous recovery of either or both of these types of associations will occur. Following A-B, C-D learning Ss were tested, at 1 of 2 retention intervals, with either a free-response recall or a stimulus-response matching task. Contrary to McGovern's (1964) findings, the data suggest that both kinds of associations show some unlearning. Alternative explanations are discussed. No evidence for spontaneous recovery was obtained. The 2nd experiment replicated the principal finding of the 1st study.

Following the study by Barnes and Underwood (1959), in which it was suggested that first-list associations are extinguished or unlearned during second-list learning, McGovern (1964) extended the analysis of unlearning to include a consideration of three types of associations which may be subject to the extinctive process. It was concluded that when specific forward (S-R) or backward (R-S) associations, or associations between the experimental context and the responses (C-R), form an A-B, A-C relationship in successive lists then extinction of the first-list associations will occur during the learning of the second list. Depending upon the paradigm, any or all of these associations will contribute to the overall decrease in first-list retention. A free-recall task, in which *S* was provided with the first-list stimuli and asked to recall the first-list responses, was assumed to reflect variations in the strengths of all three types of associations. A retention task in which the first-list stimuli and responses were provided and *S* was asked to pair the correct units was assumed to eliminate the effects of extinction of C-R associations and to reflect the combined strengths of specific forward and backward associations.

The present report contains two experiments which were prompted by this analysis of unlearning factors. The first study dealt with the A-B, C-D paradigm and asked (*a*) whether specific as well as contextual associations are extinguished during second-list learning, and (*b*) whether either or both of these kinds of associations will demonstrate spontaneous recovery. The second study attempted a partial replication of the first.

EXPERIMENT I

According to McGovern, the A-B, C-D paradigm should result in extinction of contextual but not specific first-list associations. This is because successive contextual associations in this paradigm conform in the A-B, A-C relationship while neither the S-R nor the R-S associations do so. Her data support these predictions. That is, free recall of first-list responses showed a significant decrement while matching of the first-list stimuli and responses was nearly perfect.

---

[1] This work was supported by research Grants MH 10612-01 and MH 11199-01 from the National Institute of Mental Health, Public Health Service.

There are two aspects of her situation which require comment. First, because the ability to match first-list stimuli and responses was nearly perfect (7.83 out of 8.00), there is the possibility that the specific associations were highly overlearned, that some extinction of these associations did occur, and that the matching technique was not sensitive to this reduction. Second, no matching control group was used. That is, no group was tested by the matching method after learning a single list. The only single-list control group was tested with the free-response recall method. Comparison of this control condition with an experimental group tested by the matching method is inappropriate.

Because of these factors it was thought that extinction of specific associations in the A-B, C-D paradigm should be retested. In the present study first-list learning was carried to a partial criterion to avoid the kind of ceiling effect mentioned above and both matching and response recall single-list control groups were used. In addition, two retention intervals were employed to test for spontaneous recovery. In the 2 × 2 × 2 design each of the eight groups learned either a single list (Control groups) or two lists conforming to the A-B, C-D paradigm (Experimental groups). Retention was tested either after 1 min. or 1 wk. One-half the *S*s were given the first-list stimuli and asked to recall the responses regardless of the stimuli with which they had been paired. The other half were given the first-list materials and asked to pair the correct stimuli and responses.

Method

Subjects.—The *S*s were 96 University of California undergraduates who were fulfilling an experimental participation requirement. The eight conditions were unsystematically arranged 12 times such that each occurred once in each succeeding block of eight conditions. The *S*s were assigned to this arrangement as they appeared in the testing situation.

Materials.—All lists were composed of 10 nonsense syllables as stimuli paired with 10 two-syllable adjectives as responses. Two sets of syllables and adjectives were used. One set of each was always used in the first list and the other in the second list. The Glaze (1928) meaningfulness values of the syllables ranged from 33 to 73. Within each set of syllables each vowel was used twice while no consonant was used more than once. None of the syllables in the second set began with a letter which served as a first letter in the first set. None of the syllables in the second set had more than one letter in common with any of the syllables in the first set, and if the duplication involved a consonant it occurred in a different position. The two sets of adjectives which served as responses possessed minimal inter- and intraset meaning similarity. No two adjectives began with the same letter. Two different pairings of the first-list stimuli and responses and two pairings of the second-list stimuli and responses were used equally frequently. The stimulus and response members of any given pair never began with the same letter. Three random orders of each list were constructed to minimize serial learning. Each of these orders was used equally frequently as the starting order.

Procedure.—All learning was by the anticipation method at a 2:2 sec. rate with a 4 sec. intertrial and a 1-min. interlist interval on a Stowe drum. All *S*s learned the first list to a criterion of 5 out of 10 correct responses. The four experimental groups then learned the second list for 15 anticipation trials and were tested by one of the two retention techniques after 1 min. or 1 wk. Following first-list learning the two 1-wk. control groups were dismissed with instructions to return in 1 wk. The 1-min. control groups cancelled digits for a time equivalent to the time required by the Experimental groups to learn the second list (11⅔ min.). One minute after the end of the digit-cancellation period these *S*s recalled the first list by one of the two techniques.

In the response-recall task *S*s were given a blank sheet of paper and asked to write in any order all of the first-list responses they could remember. The first-list stimuli and the entire second list were made available on cards, but *S*s were encouraged to attend to these materials only when they felt they needed them. In the matching task *S*s were given a sheet containing the first-list stimuli in a new random order. A blank space followed each of these syllables. The first-list responses were listed below these materials in a new random order. The *S*'s task was to write the correct response after each of the syllables. The entire second list was once again available to *S*s. The *S*s were allowed as much time as they wished on both tasks.

Results and Discussion

Learning.—The mean learning and recall values are contained in Table 1. An analysis of variance applied to the first-list learning data revealed no significant differences attributable to the divisions of *S*s according to the three main effects, $F\text{s} < 1$. The interactions were not significant, $F\text{s} < 2.76$. There were no significant differences among the second-list learning means, $F < 1$.

Recall.—As expected, the overall difference between the experimental and control conditions was significant, $F\ (1, 88) = 29.98, p < .01$. Recall by the Matching technique was significantly better than recall by the response-recall method, $F\ (1, 88) = 16.14, p < 0.1$. One-minute recall was superior to 1-wk. recall, $F\ (1, 88) = 62.01, p < .01$. The only significant interaction was between the recall technique and the experimental-control factor, $F\ (1, 88) = 6.01, p < .05$. This interaction reflects the fact that the difference between the Experimental and Control groups under the response-recall method was greater than the difference under the matching technique.

Since the retention interval did not interact with either of the remaining variables the 1-min. and 1-wk. data were combined in the following one-tailed comparisons. The overall control-matching mean (6.21) was significantly larger ($t = 1.7$, $p < .05$) than the overall experimental-matching mean (5.00). This significant loss in the ability to match first-list stimulus and response units in the A-B, C-D paradigm is in direct contrast to the finding reported by McGovern (1964). It suggests that some extinction of specific S-R and R-S associations may occur in this paradigm. Of course this conclusion should remain tentative because it is based upon a difference which was barely significant with a one-tailed *t* test. But if the data were found to be reliable, they would suggest that the A-B, A-C relationship need not be present for some extinction to occur. If the matching task is in fact an uncontaminated measure of specific S-R and R-S strength then there appears to have been some unlearning of specific first-list associations during second-list learning in the absence of any identifiable A-B, A-C relationship between successive S-R and R-S associations. On the other hand, there are alternative explanations of the ob-

TABLE 1

MEAN LEARNING AND RECALL VALUES

	Experimental				Control			
	Matching		Response Recall		Matching		Response Recall	
	1 Min.	1 Wk.	1 Min.	1 Wk.	1 Min.	1 Wk.	1 Min.	1 Wk.
List 1: Trials to Criterion	5.42	4.58	6.25	4.42	5.00	4.58	4.50	5.58
List 2: Number Correct	110.67	116.75	108.69	117.83	—	—	—	—
Recall: Number Correct	6.25	3.75	3.83	1.00	7.75	4.67	7.67	3.50

served decrement which would be consistent with the position that extinction is dependent upon the existence of an A-B, A-C relationship. The first of these alternatives is based upon the fact that, in the A-B, C-D paradigm, successive C-S and successive C-R associations conform to the extinctive A-B, A-C relationship. It may be that specific first-list associations are *not* affected during second-list learning but that extinction of the first-list C-S and C-R associations determine the observed decrement in the matching task. This is to say that presentation of both the stimuli and the responses during the matching task may not, as has been assumed, completely eliminate the effects of extinction of the C-S and C-R associations. The suggestion is that the specific associations may not be extinguished but that the measure of these associations is influenced by the extinction of C-S and C-R associations.

A second alternative explanation of the decrement is based upon the possibility that the unpaced-matching task does not completely eliminate competition effects and confusion between lists. Even though *S*s were allowed unlimited time, which has been assumed to minimize competition effects, complete elimination of these factors has not been demonstrated conclusively.

The overall control response-recall mean (5.59) was significantly greater ($t = 4.54$, $p < .01$) than the overall experimental response-recall mean (2.42). As noted, this difference is greater than the difference between the experimental and control matching means. If the difference between the experimental and control matching means is assumed to indicate a loss in the strength of specific associations, then this difference probably reflects a combined loss due to extinction of both specific and C-R associations.

The fact that none of the interactions involving retention interval were significant suggests that spontaneous recovery, either of the specific or the context associations, did not occur. That is, the difference between the experimental and control conditions under the two retention methods did not decrease as the retention interval increased. Percentage of RI data also failed to reveal spontaneous recovery of either type of association. The percentage of RI values for the 1-min. and 1-wk. matching conditions were both 19%. The 1-min. and 1-wk. response-recall values were 50% and 71%.

In summary, the principal finding of Exp. I, in contrast to the McGovern report, was a significant decrement in the ability to pair first-list stimuli and responses following A-B, C-D learning. It was suggested that this might indicate unlearning of specific first-list associations in the absence of the A-B, A-C relationship. Alternative explanations were discussed. The data failed to provide, in agreement with a growing number of studies (Houston, 1966; Koppenaal, 1963; Slamecka, 1966), any evidence for spontaneous recovery.

EXPERIMENT II

Method

The difference between the experimental and control matching means reported in Exp. I barely reached significance with a one-tailed test. Because of the potential importance and the lack of firm evidence for the reliability of this finding a replication of the effect was considered essential. Two of the Exp. I conditions were included in Exp. II. These were the experimental matching 1-min. and the control matching 1-min. conditions. Unlearning of specific associations would lead to superior recall in the control situation. Experiment II was conducted subsequent to the completion of Exp. I and by a different *E*. Twenty-four rather than 12 *S*s were assigned to each condition. The materials were identical, and the procedures as similar as possible, to those used in Exp.

I. The experimental and control conditions were arranged such that each occurred twice within each successive block of four *S*s. The *S*s were assigned to this arrangement as they appeared in the laboratory.

Results and Discussion

The mean numbers of first-list trials to criterion for the experimental and control groups were 7.00 and 5.83, respectively. These values did not differ significantly, $t\ (46) = 1.46$. The mean numbers of responses correctly matched in the retention task by the experimental and control groups were 6.88 and 8.25. The control mean was, as predicted and as observed in Exp. I, significantly greater than the experimental mean, $t\ (46) = 2.17$, $p < .05$. Experiment II thus offers positive evidence for the reliability of the A-B, C-D unlearning effect reported in Exp. I.

As noted, the meaning of this loss is not completely clear. It could suggest that, while *E* has not placed the learning units in an A-B, A-C relationship, *S* still experiences the "unreinforced" evocation of the first-list associations during second-list learning. For instance, while performing the second task, *S* may engage in some covert rehearsal of the A-B list. Because the experimental *S* is dealing with two sets of associations while the control *S* has learned only a single set, there may be greater confusion, and a greater tendency to perceive a given A-B association as incorrect, on the part of the experimental *S*. This explanation places a premium on the concept of unreinforced evocation and assumes that such evocations can occur in the absence of *E* determined A-B, A-C patterns. The existence of an A-B, A-C pattern may be one of the most effective determinants of unreinforced evocations but it need not be the only one.

REFERENCES

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Houston, J. P. First-list retention and time and method of recall. *J. exp. Psychol.*, 1966, **71**, 839–843.

Koppenaal, R. J. Time changes in the strengths of A-B, A-C lists; spontaneous recovery? *J. verbal Learn. verbal Behav.*, 1963, **2**, 310–319.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychol. Monogr.*, 1964, **78**(16, Whole No. 593).

Slamecka, N. J. A search for spontaneous recovery of verbal associations. *J. verbal Learn. verbal Behav.*, 1966, **5**, 205–207.

(Received June 27, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 259–263

EFFECTS OF PARTIAL REINFORCEMENT ON ACQUISITION AND EXTINCTION OF THE CONDITIONED EYEBLINK IN A MASKING SITUATION[1]

KENNETH W. SPENCE AND JOHN R. PLATT
University of Texas

The effect of partial reinforcement on the acquisition and extinction of the eyelid CR in human Ss was investigated in the context of a masking situation (probability learning). All Ss were given 60 reinforced trials and partial Ss also received an equal number of nonreinforced (delayed UCS) trials in such a manner as to equate CS, UCS, and CS-UCS presentation rates in partial and continuous groups. Extinction was carried out with an unpaired-UCS technique designed to maintain motivation. The lack of significant differences between the groups in both acquisition and extinction was interpreted as being due to the absence of the cognitive inhibitory set assumed to be operative with nonreinforcement in the standard situation. The lack of a partial-reinforcement effect in animal classical aversive conditioning studies was cited as further evidence for this interpretation.

Recent studies of the human-eyelid CR employing a masking learning task have shown that rate of extinction is much slower than in the standard conditioning situation heretofore used (Spence, 1963; Spence 1966a; Spence, Homzie, & Rutledge, 1964). Correlated with this increased resistance to extinction in the masking situation is a greatly reduced recognition on the part of *S* of the change in stimulus conditions with shift from acquisition to extinction. These findings have led the senior author (Spence, 1963; 1966b) to suggest that the very rapid extinction in the standard setup (more than half of the *S*s usually extinguish after a single nonreinforcement) reflects the effects of an inhibitory cognitive set not to respond, which *S* adopts on recognizing the shift to extinction conditions. This cognitive set is assumed not to occur in the masking situation with the consequence that the response decrement is gradual, presumably reflecting the cummulative development of a simple inhibitory process (In) with successive nonreinforced trials.

This hypothesis would interpret the slower extinction following partial than after continuous reinforcement typically found with humans in the standard conditioning situation as resulting from the absence or delayed activation of the inhibitory set in the early stages of the extinction period in the case of the partial *S*s. The reason for this delay is the much more difficult discrimination involved in going from 50% to 0% reinforcement as compared with the shift from 100% to 0% reinforcement. Presumably discrimination or recognition of the shift to zero reinforcement would be considerably delayed in partial *S*s, with the consequence that the operation of the inhibitory set would also be delayed. Hence the partial *S*s should show less early decrement in performance.

Slower extinction on the part of partial *S*s has consistently been found, in the standard set up both when the UCS was omitted on the nonreinforced trials (cf. Grant & Schipper, 1952;

---

[1] This study is part of a project concerned with motivation and set in conditioning under Contract Nonr-375(18) between the University of Texas and the Office of Naval Research. Acknowledgment is made to Abigail B. Capaldi who assisted in the collection of the data.

Hartman & Grant, 1960; Humphreys, 1939) and when the delayed UCS technique was employed (Reynolds, 1958). However, on the basis of our assumption that the inhibitory set is not operative in the masking situation, we would be led to expect no difference in the rate of extinction of partial and continuously trained *S*s in this situation. This assumes, of course, that the partial-reinforcement effect (PRE) in such classical aversive conditioning is not overdetermined, i.e., is not a function of multiple, independent factors. In the latter instance one could only expect PRE to be reduced in magnitude in the masking situation as compared with that in the standard conditioning setup.

Some evidence in support of the above reasoning is provided by an unpublished doctoral dissertation of Goldstein (1962). Employing the masking situation, Goldstein found that *S*s conditioned with a partial-reinforcement schedule (60%) involving the delayed UCS technique did not exhibit a slower rate of extinction than continuously reinforced *S*s when the extinction was carried out with the UCS omitted. However, in the case of *S*s that had a delayed UCS in extinction the performance curve of the partial *S*s did exhibit a significantly slower rate of decrement than that of the continuous *S*s.[2]

The present study is concerned with obtaining further evidence as to the conditions under which PRE occurs or does not occur in classical conditioning in humans. Use is made of a new extinction procedure employed by Spence (1966a) that maintains a high level of motivation during extinction, but which avoids the possible reinforcing effects of the delayed UCS procedure.

METHOD

Subjects.—A total of 56 students from an introductory psychology course were utilized. No *S*s were eliminated because of meeting the voluntary form response criterion suggested by Spence and Ross (1959).

Apparatus.—The *S*s were seated in a dental chair inside of a dimly illuminated IAC acoustical chamber. The CS was a 70-db. (SPL), 500-cps, 2,550-msec. tone generated by a Hewlitt-Packard 200AB oscillator and administered to *S*s through an impedance-matching network and a Lafayette trihelix speaker. An ambient noise level of 60-db. (SPL) was provided with a Foringer noise generator. The UCS was a 50-msec., 2.0-psi static air puff delivered to the cornea of the right eye through an orifice .062 in. in diameter. The equipment for recording the response and controlling the temporal relations of the stimuli was the same as that developed and used in the Iowa laboratory of the senior author (cf. Spence & Taylor, 1951). Trial programming was accomplished by means of a Western Union tape reader. The apparatus used in the probability-learning (masking) task was the same as described by Spence, Homzie, and Rutledge (1964).

Procedure.—The *S*s were first given the same instructions as were used in previous studies employing the masking task (Spence, Homzie, & Rutledge, 1964). All *S*s then received 59 reinforced trials and one nonreinforced trial at the end of acquisition, followed by 70 extinction trials. Group C, consisting of 18 men and 18 women, was conditioned on a continuous (i.e., 100%) reinforcement schedule. These *S*s received 59 paired trials with a CS-UCS interval of 500 msec. and one (last) nonreinforced trial in which the UCS was omitted. Interspersed with these reinforced trials were 60 game alone (GA) trials on which neither the CS nor UCS was present. This group thus had a total of 120 game trials on which *S* attempted to guess which light, left or right, would light up. Group P, consisting of 9 males and 11 female *S*s, received 59 reinforced trials (paired CS-UCS with ISI interval of 500 msec.) and one (last) nonreinforced trial, interspersed with 60 delayed UCS trials on which the UCS was presented 2,500 msec. after the onset of the CS and 120 game-alone trials, making a total of 240 game trials. Thus, rate of presentation of the CS and UCS was the same for both groups with the mean interval between game trials being 10 sec. and the mean interval between trials on which a CS and UCS were present being

---

[2] As will be discussed later this delayed UCS technique does not lead to any extinction following partial training in the masking situation.

20 sec. For both groups extinction consisted of 70 trials with the CS alone intermixed with 70 unpaired UCS trials. A CR was recorded whenever the recording pen showed a deflection of 1 mm. or more in the interval 150–500 msec. following the onset of the CS.

RESULTS

Performance curves in terms of percentage of CRs for the two groups are plotted in Fig. 1 for blocks of 10 reinforced trials in acquisition and blocks of 10 nonreinforced trials in extinction. Statistical analyses of these data verify the obvious lack of differences between the two groups both during acquisition and extinction. Thus the only significant result of a trend analysis of acquisition performance was the Trials effect, $F\ (5, 270) = 105.29$, $p < .001$. Likewise a trend analysis of extinction yielded Trials, $F\ (7, 378) = 79.45$, $p < .001$, as the only significant effect—all other Fs being less than 1.0. In addition the rates of extinction of the two groups were examined in terms of individual Anderson indexes (Anderson, 1963). A Mann-Whitney U test of the distribution of these indexes indicated that the extinction rates of the two groups did not differ significantly ($U = 273.5$; $Z = .47$; $p > .68$).

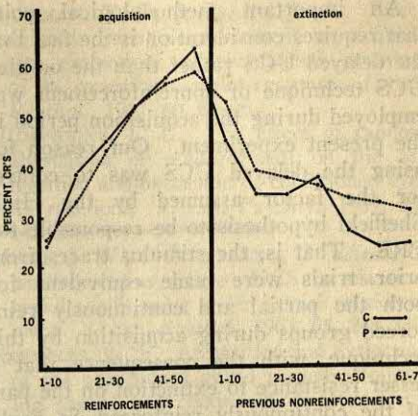


Fig. 1. Percentage of CRs in successive blocks of 10 reinforced trials (acquisition) and 10 previously nonreinforced trials (extinction).

DISCUSSION

The results of the present study add further evidence that the rate of extinction of partial and continuously reinforced *S*s does not differ in the masking situation and thus lend additional support to our hypothesis that the inhibitory set factor assumed to be responsible for the rapid extinction of continuously reinforced *S*s in standard conditioning is not present in the masking set up.

Another kind of evidence that supports this notion that PRE in classical aversive conditioning is a function of higher order cognitive processes is supplied by the findings obtained in this kind of experimental situation with animals. Quite in contrast with the consistently positive evidence for PRE in previous studies with humans, the experimental results obtained with animals have been overwhelmingly negative. These experiments have employed rats, fish, pigeons, and rabbits in a variety of aversive conditioning situations (cf. Armus, 1960; Bruner, 1965; Gonzales, Longo, & Bitterman, 1961; Gonzales, Milstein, & Bitterman, 1962; Thomas & Wagner, 1964). The few positive findings that have been obtained appear to result either from the fact that the behavior situation has involved instrumental-response components or conditions were particularly favorable for some form of stimulus aftereffects mechanism such as the Hull-Sheffield hypothesis assumes. In terms of our hypothesis these findings with animals would, of course, be interpreted as being due to the lack of higher order cognitive processes.

Also bearing on this possible relation of PRE in classical aversive conditioning to the capacity of organisms to make effective use of higher mediating activities is the finding obtained in an eyelid conditioning experiment conducted by Ross, Koski, and Yaeger (1964) with mental retardates whose IQs averaged about 19. As in the case of the animal studies these investigators found no difference in the rate of extinction of their partial and continuously trained retardates.

An important methodological point that requires consideration is the fact that the delayed UCS rather than the omitted UCS technique of nonreinforcement was employed during the acquisition period in the present experiment. Our reason for using the delayed UCS was to control for the factor assumed by the Hull-Sheffield hypothesis to be responsible for PRE. That is, the stimulus traces from prior trials were made equivalent for both the partial and continuously reinforced groups during acquisition by this technique, with the consequence that a lesser resistance to extinction on the part of the continuously reinforced *S*s could not be ascribed to a lower generalized habit strength because of a greater change in the conditioned stimulus complex in the case of these *S*s during extinction.

It should be noted, however, that the use of the delayed UCS technique during acquisition in masked conditioning led to a rather unique result, namely little or no difference in the acquisition performance of the partial and continuously reinforced groups when compared in terms of equated number of reinforced trials, i.e., trials on which CS and UCS were paired at the 500-msec. interval. Such a result, which Goldstein (1962) also obtained in two independent experiments, differs markedly from the acquisition data obtained with human *S*s in previous standard experiments involving partial reinforcement. Thus the latter have found that the performance level of a group trained with 50% reinforced trials falls significantly below that of a 100% reinforced group. In terms of equated reinforced trials these studies, whether they employed the omitted UCS (Froseth & Grant, 1961; Grant & Schipper, 1952; Hartman & Grant, 1960) or the delayed UCS (Ross, 1959; Spence & Trapold, 1961), have typically found the performance level of the 50% group to be between 55 and 70% of that of the 100% group.[3] In the case of Reynold's (1958) study, in which the partial group was reinforced on 60% of the trials, this percentage was somewhat higher (75%). In contrast to these values for the standard set up, comparably determined values for the masking situation (present study and two Goldstein experiments) approach, of course, 100%, as there is little or no difference between the partial and continuously reinforced groups.

It is apparent from these findings that the effects of a delayed UCS trial in the masking situation is markedly different from that which it has in the standard situation. In the masking situation a delayed UCS trial appears to be entirely neutral in its effects, leading neither to an increment nor a decrement in CR frequency, at least as far as group data are concerned. Further confirmation of this neutral effect of the delayed UCS is the finding of Goldstein (1962) that the use of 100% delayed UCS trials in an "extinction" series following partial training that had involved 60% delayed UCS trials, led to little if any decrease in the frequency of CRs. Thus one experiment gave a decrease of 3% from the final acquisition level to Trials 71–80 of the extinction series, while the decrease in a second experiment was 11% over 80 trials. This decrease occurred in the first 10 trials after which the curve fluctuated around a constant level of performance.

Finally, attention is directed to the fact that the occurrence of PRE in the standard situation is primarily a function of the extremely rapid extinction of the continuously reinforced group that results, in terms of our hypothesis, from the immediate action of an inhibitory set. Partial reinforcement is assumed to delay the action of this cognitive set and thus retard response decrement. The absence of PRE and the even slower extinction of both groups than in the standard experiment is taken as evidence of the absence of a cognitive factor in the masking situation. However, it should be pointed out that the use of the delayed UCS procedure does not provide for *nonreinforcement* effects as it does in the acquisition period in the standard experiment. Whether, for example, the technique of omitting the UCS in the context of the masking task would result in both lower acquisition performance and greater resistance to extinction for such partial as compared with continuously reinforced *S*s remains to be investigated.

---

[3] The evidence of these studies suggests that the decremental effects of partial (50%) reinforcement in the standard type of experiment is about the same for the two methods of nonreinforcement, omitting or delaying the UCS. Thus the mean of the estimated percentage values for the three studies conducted with the omitted UCS was 62%, while the delayed UCS studies gave a value of 61%. Differing sharply with the implication of these findings that there is little or no difference in the performance of *S*s that had these two nonreinforcement procedures is the finding of Moore and Gormezano (1963) which showed that a group with 50% omitted UCS trials responded at a much higher absolute level (85%) than a group that had 50% delayed UCS trials (59%) in acquisition.

REFERENCES

ANDERSON, N. H. Comparison of different populations: Resistance to extinction and transfer. *Psychol. Rev.*, 1963, **70**, 162–179.

ARMUS, H. L. Effect of percentage of reinforcement and distribution of acquisition trials on extinction of conditioned fear. *Psychol. Rep.*, 1960, **11**, 387–390.

BRUNER, A. UCS properties in classical conditioning of the albino rabbit's nictitating membrane response. *J. exp. Psychol.*, 1965, **69**, 186–192.

FROSETH, J. Z., & GRANT, D. A. Influence of intermittent reinforcement upon acquisition, extinction, and spontaneous recovery in eyelid conditioning with fixed acquisition series. *J. gen. Psychol.*, 1961, **64**, 225–232.

GOLDSTEIN, H. The effects of set factors and reinforcement schedules on the acquisition and extinction of the conditioned eyelid response. Unpublished doctoral dissertation, State University of Iowa, 1962.

GONZALES, R. C., LONGO, N., & BITTERMAN, M. E. Classical conditioning in the fish: Exploratory studies of partial reinforcement. *J. comp. physiol. Psychol.*, 1961, **54**, 452–456.

GONZALES, R., MILSTEIN, S., & BITTERMAN, M. E. Classical conditioning in the fish: Further studies in partial reinforcement. *Amer. J. Psychol.*, 1962, **75**, 421–428.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, **60**, 89–96.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

MOORE, J. W., & GORMEZANO, I. Effects of omitted versus delayed UCS on classical eyelid conditioning under partial reinforcement. *J. exp. Psychol.*, 1963, **65**, 248–257.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., KOSKI, C. H., & YAEGER, J. Classical eyelid conditioning of the severely retarded: Partial reinforcement effects. *Psychon. Sci.*, 1964, **1**, 253–254.

SPENCE, K. W. Cognitive factors in the extinction of the conditioned eyelid response in humans. *Science*, 1963, **140**, 1224–1225.

SPENCE, K. W. Extinction of the human eyelid CR as a function of presence or absence of the UCS during extinction. *J. exp. Psychol.*, 1966, **71**, 642–648. (a)

SPENCE, K. W. Cognitive and drive factors in the extinction of the conditioned eyeblink. *Psychol. Rev.*, 1966, **73**, 393–480. (b)

SPENCE, K. W., HOMZIE, M. J., & RUTLEDGE, E. F. Extinction of the human eyelid CR as a function of the discriminability of the change from acquisition to extinction. *J. exp. Psychol.*, 1964, **68**, 545–552.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determinants of amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & TRAPOLD, M. A. Performance in eyelid conditioning as a function of reinforcement schedules and changes in them. *Proc. Nat. Acad. Sci.*, 1961, **47**, 1860–1868.

THOMAS, E., & WAGNER, A. R. Partial reinforcement of the classically conditioned eyelid responses in the rabbit. *J. comp. physiol. Psychol.*, 1964, **58**, 157–158.

(Received June 30, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 264–271

MODES AND ADEQUACY OF RESOLUTION OF THREE BASIC TYPES OF COGNITIVE-MOTOR CONFLICT[1]

SEYMOUR EPSTEIN[2] AND BARRY SMITH

University of Massachusetts

In an oft-cited study, Hovland and Sears reported blocking to be the most frequent response to avoidance-avoidance conflict. As their instructions did not require S to leave the starting point, blocking, as they defined it, was the most adequate response possible. In the present study, in which remaining at the starting point was defined as a poor response, the incidence of blocking was markedly reduced, and the most prevalent response to avoidance-avoidance conflict was found to be a single response to 1 of the negative goals. A comparison of the 3 basic types of conflict on quality and speed of response indicated that the conflicts could be ordered in increasing difficulty as follows: approach-approach, approach-avoidance, and avoidance-avoidance conflict.

In an oft-cited series of studies, Hovland and Sears (1938), and Sears and Hovland (1941) investigated the four basic kinds of conflict described by Lewin (1935). The *S*s were instructed to move a pencil as quickly as possible across a paper to the corner in which a green light appeared, and away from the corner in which a red light appeared. After a number of single trials, conflict trials were presented by flashing more than one light. For approach-approach (Type I) conflict, a green light appeared in each corner, for avoidance-avoidance (Type III) conflict, a red light appeared in each corner, for approach-avoidance (Type II) conflict, a green and red light appeared in the same corner, and for double approach-avoidance (Type IV) conflict, green and red lights appeared in both corners. They reported that blocking was the predominant response to Type III conflict, and that blocking increased as strengths of the conflicting responses approached equality. While the major findings in both studies concerned blocking, the experimental procedure failed to provide an unambiguous measure of blocking. Blocking, as defined, was the most adequate response possible, as it maximized the distance from both avoidance goals. To an unknown extent, the group that exhibited blocking could thus be assumed to be composed of some *S*s who behaved in a rational manner, and others who froze.

In a recent study, Ringuette (1965) not only followed the procedure of Hovland and Sears in treatment of the blocking response, but pooled responses to different conflicts without regard to adequacy of resolution. For example, all double responses were designated as "equivocation," whether they were appropriate responses to Type I conflict or poor responses to Type III conflict. A major aim of the present study was to evaluate modes of conflict resolution with respect to adequacy for each of the conflict types.

A second aim of the study was to investigate difficulty of conflict resolu-

---

[1] This paper was supported by Grant MH 01293 from the National Institute of Mental Health to S. Epstein.

[2] The paper was written while Epstein was a visiting professor at the Department of Social Relations, Harvard University.

tion by the use of measures of time, speed, errors, and detours. By the introduction of a simple maze we were able to increase response times to the point where they could be reliably measured on single trials, and to obtain data on errors, in the form of crossing the boundaries of the maze, and on detour responses.

The two basic assumptions from which hypotheses were formulated were as follows: (*a*) Conflicts tend to be resolved with the most appropriate responses possible under the circumstances provided; and (*b*) the basic types of conflict vary in the degree to which they disrupt cognitive functioning. The second assumption obviously interacts with the first. Whether any of the conflicts in a particular situation is of sufficient severity to cause predominantly maladaptive responses cannot be determined on an a priori basis. Thus, hypotheses on the predominant response to each type of conflict were formulated with respect to Assumption *a*, only. Hypothesis 4 indicates the order in which Assumption *b* can be expected to interfere with Assumption *a*.

Hypothesis 1: Type I conflict elicits predominantly single and double responses.

Hypothesis 2: Type II conflict, when the avoidance is stronger than the approach incentive at the goal, elicits predominantly single responses to the opposite corner of the conflicted goal.

Hypothesis 3: Type III conflict elicits predominantly compromise responses which maximize the distances from negative goals.

Hypothesis 4: Inadequacy of response, whether measured by mode of resolution, time, speed, or errors, increases in the following order of conflict types: I, II, III.

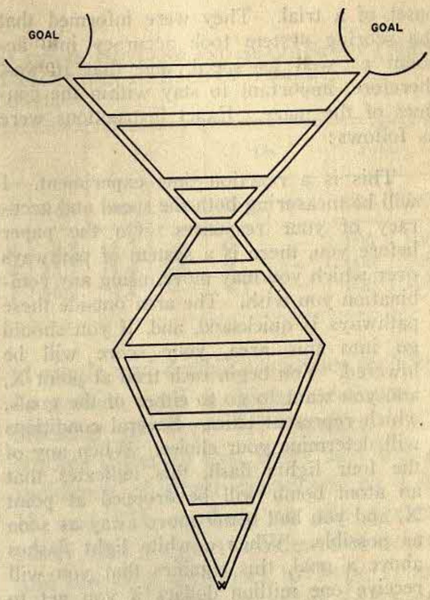


FIG. 1. Maze which was used for all trials.

METHOD

Subjects.—The *S*s were 42 male and 28 female students in an introductory psychology class. They were randomly assigned to groups counterbalanced for order and sequence.

Apparatus.—The apparatus consisted of a masonite board, 12 × 16 in., in the center of which was a recessed area which contained a stack of 8½ × 11 in. sheets of paper on each of which was printed the same maze. At the upper right and left corners of the board were a pair of red and white lights. The lights were operated by *E* with the aid of a control panel which was screened from *S*'s view by a masonite board mounted on the back of the conflict apparatus. Response time was measured to the nearest .01 second with a Standard Electric timer. The maze is presented in Fig. 1. It was constructed so that the first section of each side directed *S*'s vision toward one goal, the second toward the opposite goal, and the third toward the first goal. Crossways permitted detour responses from one side of the maze to the other.

Procedure.—Each *S* was run individually. The *S*s were instructed (*a*) to approach or avoid a goal as a function of which light was presented, (*b*) to stay within the pathway, and (*c*) to leave the starting point upon the onset of a trial. They were informed that the scoring system took accuracy into account as well as speed, and that it was, therefore, important to stay within the confines of the maze. Exact instructions were as follows:

> This is a reaction time experiment. I will be measuring both the speed and accuracy of your responses. On the paper before you, there is a system of pathways over which you may move, using any combination you wish. The area outside these pathways is quicksand, and, if you should go into this area, your score will be lowered. You begin each trial at point X, and you want to go to either of the goals, which represent cities. Several conditions will determine your choice. When any of the four lights flash, this indicates that an atom bomb will be dropped at point X, and you had better move away as soon as possible. When a white light flashes above a goal, this signifies that you will receive one million dollars if you get to that city quickly enough. If a red light flashes above a goal, this signifies that an atom bomb will be dropped at that city.
>
> In summary, when any light goes on, you must get away from point X. You are to go toward a white light and away from a red light. Any questions? Remember to move as quickly as possible while remaining within the boundaries.[3]

There were 43 trials in all. Of these, Trials 21, 32, and 43 were conflict trials, and the others single approach or avoidance trials. To begin a trial, *S* placed his pencil at the starting point (X). The *E* then simultaneously activated the lights according to a prearranged design and started the electric timer. Time was stopped when *S* finished moving the pencil and said "Stop," indicating he had, in fact, completed his response. Response time was recorded and the next trial administered.

Design.—Each *S* received all three types of conflict. Sequence of presentation was permuted according to a 3 × 3 Latin square. All single, or nonconflict trials, were received in the same order by all *S*s and were randomized and equal with respect to frequency of approach and avoidance, and position.

RESULTS

Mode of conflict resolution.—Six categories were employed in classifying mode of conflict resolution, four of which were the same as those used by Hovland and Sears (1938), and two of which were specific to the use of a maze. The six modes were as follows, with a subscore for an adequate resolution designated by a plus sign, and an inadequate resolution by a minus sign:

1. Single response (S+ or S−). The goal is reached directly, without detour. An S+ is scored for an approach response to a positive goal and an avoidance response to a negative goal. An S− is scored when approach is to a negative goal or away from a positive goal. In Type I conflict, all single responses are +; in Type II conflict, a single response away from the conflicted goal is +, and to the conflicted goal is −; in Type III conflict, all single responses are −.

2. Double response (Db+ or Db−). Both goals are reached by either going from one goal directly to the other, or by returning to the starting point and then proceeding to the other goal. A Db+ can only be obtained in Type I conflict.

3. Compromise (C+ or C−). The response does not terminate at a goal, but at some point on the pathway of the maze between the starting point and a goal point. A C+ could be obtained on Type III conflict by terminating within the middle one-third of the maze, and thereby maximizing the distance from the three negative points. All other compromise responses were scored −.

4. Detour (Dt−). The response proceeds through one or more cross-

---

[3] The euphemistic reference to atomic bombs, a million dollars, and quicksand was for the purpose of increasing interest and having *S*s imagine strong incentives. That the results were not specific to the imaginary incentives has since been confirmed in further work which has systematically varied incentive.

TABLE 1

PERCENTAGE OF CASES IN PRESENT STUDY (E&S) AND IN HOVLAND AND SEARS STUDY (H&S) RESPONDING WITH EACH MODE OF RESOLUTION TO EACH OF THREE TYPES OF CONFLICT

Type of Conflict	S		Db		C		B		Dt		Ds		Total	
	E&S	H&S	E&S	H&S	E&S	H&S	E&S	H&S	E&S	H&S	E&S	H&S	E&S	H&S
I	62.9	57.5	27.1	21.2	2.9	12.5	1.4	8.8	5.7	—	0	—	100	100
II	71.5	14.4	7.1	46.9	8.6	10.6	1.4	28.1	11.4	—	0	—	100	100
III	47.1	17.5	1.4	7.5	31.3	28.8	14.7	46.2	1.4	—	4.1	—	100	100

ways. Detour responses were always −. A response to Type III conflict that used a crossway to remain within the middle third of the maze was not considered a detour response, and was scored as a C+.

5. Disorganized (Ds−). This was a response in which *S* made no attempt to remain within the confines of the maze.

6. Blocking (Bl−). This was a failure to leave the starting point following stimulus onset.

Table 1 presents the frequency of the different modes of conflict resolution in the present study and in the Hovland and Sears (1938) study. In both studies there is a high incidence of single and double responses to Type I conflict, and a high incidence of blocking to Type III conflict. The studies differ in that there is a higher incidence of single responses to Type II and III conflicts in the present study, and a higher incidence of double and blocking responses to Type II and III conflicts in the Hovland and Sears study.

Table 2 presents a complete breakdown of responses according to type of conflict and mode and adequacy of resolution. Nonconflict control trials are represented by the control trials immediately preceding the three conflict trials, divided by three, and rounded to the nearest whole number.

TABLE 2

MODE AND ADEQUACY OF RESPONSE AS A FUNCTION OF CONFLICT TYPE

MODE	CONFLICT TYPE			
	Nonconflict	Type I	Type II	Type III
S+	66 (94.3%)	44 (62.9%)	31 (44.3%)	—
S−	2 (2.9%)	—	19 (27.2%)	33 (47.1%)
S_{tot}	68 (97.1%)	44 (62.9%)	50 (71.5%)	33 (47.1%)
Db+	—	19 (27.1%)	—	—
Db−	0 (0.0%)	—	5 (7.1%)	1 (1.4%)
Db_{tot}	0 (0.0%)	19 (27.1%)	5 (7.1%)	1 (1.4%)
C+	—	—	—	20 (28.6%)
C−	0 (0.0%)	2 (2.9%)	6 (8.6%)	2 (2.7%)
C_{tot}	0 (0.0%)	2 (2.9%)	6 (8.6%)	22 (31.3%)
B−	0 (0.0%)	1 (1.4%)	1 (1.4%)	10 (14.7%)
Dt−	2 (2.9%)	4 (5.7%)	8 (11.4%)	1 (1.4%)
Ds−	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (4.1%)
Total +	66 (94.3%)	63 (90.0%)	31 (44.3%)	20 (28.6%)
Total −	4 (5.7%)	7 (10.0%)	39 (55.7%)	50 (71.4%)
TOTAL	70 (100%)	70 (100%)	70 (100%)	70 (100%)

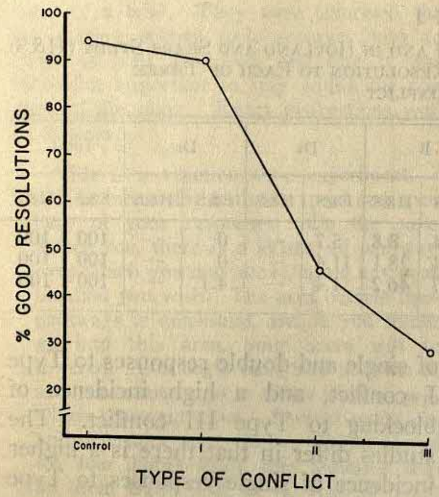


FIG. 2. Percentage of good solutions as a function of conflict type.

It is apparent from Table 2 and Fig. 2 that there is a decrease in the number of good solutions in the order: nonconflict control trials, Type I conflict, Type II conflict, and Type III conflict. The differences between the control trials and Type I conflict and between Type II and Type III conflict are small. By far the largest consecutive difference occurs between Type I and Type II conflict. It is of interest that even on control trials there were 5.7% errors, which included two single responses and two detour responses. The set for speed apparently produced poor solutions even on nonconflict trials, indicating the necessity for taking into account reactions to nonconflict trials when evaluating conflict trials.

The number of *S*s who made poor responses to each of the different types of conflict was compared with the number who made poor responses on control trials. To obtain a chi-square uncontaminated by dependent measures, data were evaluated using only *S*s who performed differently on conflict and nonconflict trials, and comparing the number who made more errors on conflict than control trials and the number who made errors in the reverse direction, with an expectancy of an even division. Chi-square was not significant for approach-approach conflict, but the other two types of conflict were significant at the .001 level. Using the same procedure for comparisons among conflict types, Types I and II and Types I and III conflict differed significantly from each other at the .001 level, and Types II and III conflict at the .05 level. In sum, Types II and III, but not Type I, conflict produced more poor solutions than their nonconflict control trials. Among the conflict types, Type III produced the most and Type I the least poor solutions.

An examination of poor solutions by conflict type (see Table 2) indicated that the most frequent poor solution to Type II conflict was a single response to the conflicted goal, and that Type II conflict also elicited the greatest number of inappropriate detour and compromise responses. Type III conflict produced the greatest number of poor single responses and also the greatest incidence of blocking. Type I conflict elicited too few poor responses to warrant serious consideration of any further breakdown.

Time and error scores.—Time was scored from stimulus onset to the point where *S* completed the response and called time. Ten *S*s were eliminated because of a failure to call time on one or more critical trials, leaving a total *N* of 60. An analysis of variance was first performed on the single approach and avoidance trials to determine whether latency varied as a function of moving toward vs. moving away from a goal. The 40 nonconflict trials were divided into four blocks of 10 trials each. Scores were represented by the median response within a block. Thus, there were four scores per *S*, which provided a measure of change due to practice. No significant differences were found for approach vs. avoidance, for practice effects, or for their interaction.

An analysis of variance was next performed on the conflict trials. Scores consisted of response time for the conflict trial minus the median response time for the three practice trials that immediately preceded it. The only significant source of variance was conflict type, which was significant at the .001 level. Figure 3 presents the data in absolute time scores. An analysis of contrasts, with the *F* for each contrast evaluated against the conservative *F* recommended by Scheffé (1959) for multiple comparisons, revealed that: (*a*) Type II and III conflicts produced significantly longer response times than Type I conflict ($p < .005$), and (*b*) there was no difference between Type II and III conflicts ($F = .008$).

A problem that arises from the use of time as a score is that the different modes of conflict resolution differ in average amount of distance traversed. Thus, the reason Type III conflict failed to elicit higher response times than Type II conflict could have been that fewer single and double responses were produced to Type III conflict, so that the average distance traveled was less. To control for distance, an analysis was performed on speed of response, obtained by dividing distance by time. Conflict type was significant at the .05 level. In Fig. 4, it can be seen that there is a continuous decrease in speed in the order: nonconflict control trials, Type I conflict, Type II conflict, and Type III conflict.

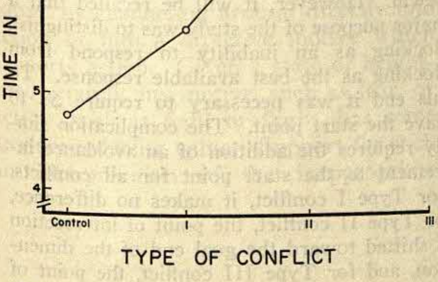


FIG. 3. Mean response time as a function of conflict type.

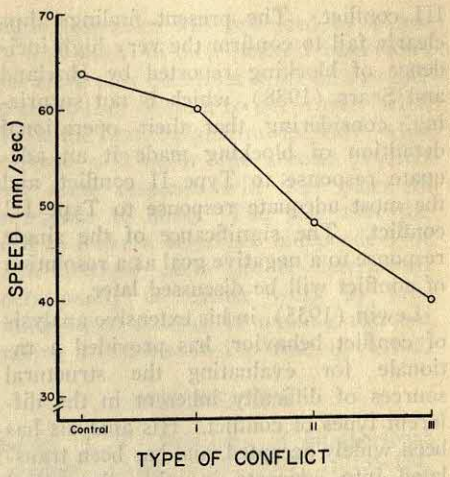


FIG. 4. Speed of response as a function of conflict type.

It was not possible to analyze error scores, as the number of errors, defined by the amount of movement outside the pathways, was negligible.

DISCUSSION

The hypothesis that adaptive modes of conflict resolution predominate was supported for Type I and II, but not Type III conflict. The hypothesis that Type III conflict is the most, and Type I conflict the least, disruptive was confirmed.

While blocking occurred more often to Type III than to any other conflict, it was a much less frequent response than the single response to both Type II and Type

III conflict. The present findings thus clearly fail to confirm the very high incidence of blocking reported by Hovland and Sears (1938), which is not surprising, considering that their operational definition of blocking made it an adequate response to Type II conflict, and the most adequate response to Type III conflict. The significance of the single response to a negative goal as a resolution of conflict will be discussed later.

Lewin (1935), in his extensive analysis of conflict behavior, has provided a rationale for evaluating the structural sources of difficulty inherent in the different types of conflict. His analysis has been widely accepted, and has been translated into concepts in other theoretical systems (Brown, 1957; Hull, 1938; Miller, 1944). It is thus of interest to examine the present findings with reference to Lewin's formulations. According to Lewin, Type I conflict produces a state of unstable equilibrium, and therefore is apt not to be disruptive, although it may produce an initial delay until the first movement occurs toward one goal or the other. The data support Lewin's hypothesis. Mode of conflict resolution was not significantly different for control trials and Type I conflict, although response time was slightly greater for the latter. That the increase in response time was not merely a consequence of the double responses that were made to Type I conflict was verified by a further analysis restricted to single responses.

Type II, or approach-avoidance, conflict is of particular interest, as when gradients of approach and avoidance are of appropriate strength, a presumably steeper gradient of avoidance, by intersecting a gradient of approach, produces a state of stable equilibrium. It follows that an individual with such a conflict can neither leave the field nor obtain the desired goal, but is trapped at an intermediate point by his conflicting desires (Lewin, 1935; Miller, 1944). In the present study, the avoidance reaction at the goal should be stronger than the approach reaction, as death should outweigh a million dollars. Since the starting point in the present experiment is negative, responses should fall somewhere between the start point and the conflicted goal.[4] The same prediction holds if the gradients cross or if avoidance is greater than approach at all points; the predominant response to Type II conflict must terminate somewhere other than at the conflicted goal or the starting point. The predicted responses include compromise responses, detours, and single responses to the nonconflicted goal. Single responses to the conflict goal should be rare. While the most frequent response was an appropriate single response, which is in support of the model, the second most frequent response (27% of all responses) was a single response to the conflicted goal, which is not in support of the model. Turning to the cognitive characteristics of Type II conflict, it is evident that Type II requires a more difficult discrimination than Type I conflict, as only one of the two corners is correct. The presentation of opposing stimuli within the same location can be expected to cause confusion and error by making it difficult to disregard the inappropriate element. The high incidence of poor single responses will be discussed under Type III conflict, where they occurred with yet greater frequency.

Type III, or avoidance-avoidance, conflict according to the Lewinian model should elicit, predominantly, responses that fall between the two avoidance goals and the negative starting point. While theoretically compromise responses should be most frequent, it was found that the most frequent response, by far, was the single response. Compromise responses were second, and blocking third. The high incidence of inappropriate single responses to Type II and III conflicts violates the Lewinian model. An analysis of Type III conflict, as presented in the present study, required three points, two negative goals and the starting point, to be taken into account. Moreover, the response had to be terminated in a vague region rather than at a specific goal. While it is clear why Type III was cognitively the most difficult conflict, it is not apparent why errors consisted primarily of poor single responses and not other responses, such as blocking. Two explanations, not incompatible, can account for the high incidence of single responses to the negative goal. One follows from the consideration that the single response was the most practiced response, and so had the highest habit strength. Given a state of momentary heightened drive produced by conflict, the response with the highest habit strength could be expected to become increasingly prepotent. This explanation is consistent with Brown and Farber's (1951) analysis of conflict as producing a state of heightened drive.

---

[4] Making the start point negative complicates the models of conflict described by Lewin. However, it will be recalled that a major purpose of the study was to distinguish blocking as an inability to respond from blocking as the best available response. To this end it was necessary to require *S*s to leave the start point. The complication simply requires the addition of an avoidance increment to the start point for all conflicts. For Type I conflict, it makes no difference, for Type II conflict, the point of intersection is shifted toward the goal end of the dimension, and for Type III conflict, the point of resolution is shifted to a point some distance from, rather than directly at, the start point.

The second explanation is suggested by an analysis of anxiety as distinct from fear and physiological arousal (Epstein, in press). Briefly, anxiety is viewed as a noxious state produced by a heightened degree of *undirected* physiological arousal following the perception of danger. Putting it otherwise, anxiety is conceived of as unresolved fear. Anxiety, which is more disturbing than fear, can be resolved by channeling arousal directly into action or into a directive motive which supports action. Given a state of heightened arousal, any motive, such as fear, or action, such as fight or flight, has some anxiety-reducing value. Support for such an interpretation of the inappropriate single responses was provided by a few *S*s who stated in their posttest interviews that since they could not think of a proper solution, they decided to get it over with. While the first explanation indicates what response will occur, the second predicts that some response will take place in preference to none. The role that the need to make a response, independent of habit strength, plays can be determined operationally in further work by providing equal practice in making and not making a response in the training period.

REFERENCES

Brown, J. S. Principles of intrapersonal conflict. *Confl. Resol.,* 1957, **1**, 135–154.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables with suggestions toward a theory of frustration. *Psychol. Bull.,* 1951, **48**, 465–495.

Epstein, S. Toward a unified theory of anxiety. In B. Maher (Ed.), *Progress in experimental personality research,* Vol. IV, in press.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exp. Psychol.,* 1938, **23**, 477–493.

Hull, C. L. The goal-gradient hypothesis applied to some "field-force" problems in the behavior of young children. *Psychol. Rev.,* 1938, **45**, 271–299.

Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1 New York: Ronald Press, 1944. Pp. 431–465.

Ringuette, E. L. Selected personality correlates of mode of conflict resolution. *J. pers. soc. Psychol.,* 1965, **2**, 506–512.

Scheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Sears, R. R., & Hovland, C. I. Experiments on motor conflict: II. Determination of mode of resolution by comparative strength of conflicting responses. *J. exp. Psychol.,* 1941, **28**, 280–286.

(Received June 30, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 272–281

FORMATION, MAINTENANCE, GENERALIZATION, AND RETENTION OF RESPONSE HIERARCHIES: THE ROLE OF MEANINGFULNESS OF RESPONSE MEMBERS[1]

ALBERT E. GOSS[2] AND NANCY J. COBB[3]

University of Massachusetts

Lists of 4 divergent paired-associate units were presented under relative percentages of occurrence of alternative response members (RP) of 50-50, 60-40, 70-30, 80-20, and 90-10. The alternative response members of all units of a list were both of high (HH) or low (LL) meaningfulness, or of high meaningfulness and relatively more (HL) or less (LH) frequent than alternatives of low meaningfulness. Under RP from 60-40 upward, rates of acquisition of anticipations of more frequent response members and also percentages of those responses during the last 20 trials and during trials with stimulus members alone in the same and different contexts were typically higher with HL and HH lists than with LL and LH lists. Under RP of 50-50, rates and percentages were highest for high meaningfulness members of HL, LH lists, intermediate with HH and LL lists, and lowest for low meaningfulness members of HL, LH lists.

A response hierarchy can be defined as the occurrence of two or more responses in alternation to successive presentations of a stimulus. Thus defined, the present experiment investigated acquisition or formation, maintenance, generalization, and retention of response hierarchies as functions of relative percentages of occurrence of alternative response members (RP) and of meaningfulness of those members. Divergent paired-associates (PA) units were used to form the response hierarchies. Each unit consisted of a stimulus member which, on $P\%$ of its presentations, was accompanied by one of two response members, and which, on $(1-P)\%$ of its presentations, was accompanied by the other response member.

Maintenance of hierarchies was tested by presenting stimulus members alone in the apparatus used for acquisition of responses with the divergent PA units (same [S] context). Generalization of hierarchies was tested by presenting stimulus members alone in a context different from that of formation (different [D] context). Retention of hierarchies was assessed by presenting the stimulus members alone in the S and D contexts both 2 sec. or immediately after and 24 hr. after the formation trials.

In previous experiments of this series (Goss, 1965; Goss & Cobb, 1966; Goss & Sugerman, 1961), the alternative response members of each unit were both of high meaningfulness. Added for the present experiment was a list in which both response members of each unit were of low meaningfulness. Also added were lists in which one response member of each unit was of high meaningfulness and the other was of low meaningfulness. The lists in which both response members of each unit were of high meaningfulness or low meaningfulness are designated HH and LL, respectively. In the lists in which one response member of each unit was of high meaningfulness and the other of low meaningfulness, response members of high meaningfulness occurred more frequently or less frequently. These two assignments of percentages of occurrence of alternative response members are designated HL and LH, respectively.

---

[1] The present experiment is one of a series on paired-associates learning under variations in occurrence of stimuli in a program of research sponsored by the Office of Naval Research through Contract 3357(03) with the University of Massachusetts. Nancy Farrick Feldman aided in gathering and analyzing the data. Helen L. Rowell prepared the materials, contributed to analyses of the data, and aided in the completion of this report.

[2] Now at Douglass College, Rutgers University.

[3] Now at Los Angeles State College.

By presenting these lists under RP of 50-50, 60-40, 70-30, 80-20, and 90-10, effects of meaningfulness of response members under different relative percentages of their occurrence could be compared. Under RP of 50-50, the expectation was of a more rapid increase in percentages of anticipation of response members of high meaningfulness than of those of low meaningfulness. Under RP of 60-40 and above, the expectation was of more rapid increases in percentages of anticipation of the more frequent response members of the lists in an order HL > HH > LL > LH. Of interest, too, was whether or not these expected differences in rate of acquisition would be accompanied by like differences in apparent final or asymptotic percentages of anticipation and of recall.[4]

One general basis for the expectation of acquisition rates in the order HL > HH > LL >LH was findings of a direct relationship between acquisition rate and meaningfulness of response members with PA units in a 1:1 pattern of pairings of stimulus members and response members (Goss & Nodine, 1965). Another general basis was findings of a direct relationship between percentages of anticipation of more frequent response members and their actual percentages of occurrence (Goss, 1965; Goss & Cobb, 1966; Goss & Sugerman, 1961). With the HL list, the more frequent response members have the advantages of greater relative frequency and of greater meaningfulness. With the HH list, these members have the advantage of greater relative frequency. With the LL list, the more frequent response members have the disadvantage of low meaningfulness relative to those members in the HL and HH lists, but the advantage of greater relative frequency than the alternative response members of equal low meaningfulness. With the LH list, the more frequent response members have the disadvantages of lower meaningfulness both relative to those members in the HL and HH lists and relative to the alternative less frequent response members.

With PA units in a 1:1 pattern, anticipations of or recall of response members of different meaningfulness typically reach or would seemingly reach the same final level. With PA units in a 1:2 pattern of pairings or, more generally, in a 1:*n* pattern, percentages of anticipation and recall of the more frequent response members of the different lists might stabilize at the same level. But they might stabilize at different levels, perhaps in the HL > HH > LL > LH order expected for differences in acquisition rates.

METHOD

Lists and apparatus.—Shown in Table 1 are the CVC stimulus members and the pairs

---

[4] Anticipation refers to *S*'s responses during acquisition by means of an anticipation format with response members present. Recall refers to *S*'s responses during trials with response members omitted.

TABLE 1

LISTS WITH RESPONSE MEMBERS BOTH OF HIGH OR LOW MEANINGFULNESS OR ONE OF HIGH AND THE OTHER OF LOW MEANINGFULNESS

Stimulus Members	Response Members					
	HH		LL		HL and LH[a]	
					List A	List B
FES (69)	VIC (97)	BAL (91)	WUC (16)	GAQ (11)	BAL WUC	VIC GAQ
NAV (72)	HOB (90)	SUR (91)	XOP (7)	MUJ (10)	SUR XOP	HOB MUJ
RUF (86)	CEN (75)	DOL (94)	KIJ (8)	ZOS (13)	DOL KIJ	CEN ZOS
WOM (67)	TEX (100)	LIK (91)	YEF (11)	QIH (6)	LIK YEF	TEX QIH

Note.—Archer values are in parentheses.

[a] With response members of high meaningfulness more frequent, HL; with response members of low meaningfulness more frequent, LH.

of CVC response members of each of the four units of the lists. The units of the HH list are those employed previously (Goss, 1965; Goss & Cobb, 1966; Goss & Sugerman, 1961). Archer (1960) association values of the stimuli of this and of the other lists are in parentheses. The stimulus members common to all four lists are of relatively high meaningfulness and of low formal similarity to each other and to the response members of each list. The response members of each list are of low formal similarity to each other.

These CVCs were typed in pica capitals on tapes for presentation on a Stowe memory drum. Details of the randomization of occurrences of stimulus members, regardless of occurrences of response members, can be found in Goss and Sugerman. Under RP of 50-50, occurrences of alternative response members of each unit within successive blocks of 10 trials were in terms of different, randomly selected Gellerman orders (Hilgard, 1951). For RP of 60-40, one of the five occurrences of one of the response members of each unit under RP of 50-50 was randomly changed to the occurrence of the other response member. Remaining occurrences of one of the response members of each unit were successively changed one at a time to occurrences of the other response member to realize RP of 70-30, 80-20, and 90-10.

With HH and LL lists, occurrences of the more frequent and less frequent response members under RP of 60-40 and above were counterbalanced. In all lists, response members of the four units occurred at the same RP; thus RP varied "between" rather than "within" lists.

Procedure.—In previous experiments with the HH list, anticipations of response members were essentially stable by Trials 61–80. Therefore, with this list only 80 trials were administered. With the other three lists, 120 trials were administered. The units were presented at a 2:2-sec. rate. As tapes were changed after every 40 trials, breaks of 10–20 sec. occurred. The instructions resembled those standard for PA learning with conventional units of one response member; anticipations were spelled.

Either about 2 sec. after or 24 hr. after the last block of 20 formation or acquisition trials, 20 additional trials were administered with the stimulus members alone (response members omitted). They were presented at a 2-sec. rate. No information was given *S*s regarding omission of the response members. Any inquiries as to what to do under the changed condition were answered by instructions to respond as previously.

Upon completion of the 20 trials with stimulus members alone in the S context, *S*s were given 20 trials with stimulus members alone in the D context. Specifically, the context was a sheet of paper with the stimuli printed in pica capitals in four columns of 20 each. Instructed "to write any word or syllable you think of to the right of each stimulus," *S*s proceeded at their own pace.

Subjects.—First at the immediate retention interval and then at the 24-hr. interval, the 20 combinations of RP and lists were administered in counterbalanced cycles. At each interval, males and females drawn from the courses in introductory psychology were assigned to particular combinations in order of appearance until an *N* of eight was reached. Thus, data are reported on 320 *S*s. Sex of *S*s was controlled but not treated as a variable.

TABLE 2

MEANS OF PERCENTAGES OF ANTICIPATION OR RECALL OF MORE FREQUENT AND LESS FREQUENT RESPONSE MEMBERS OF EACH OF THE LISTS DURING SUCCESSIVE BLOCKS OF 20 ACQUISITION TRIALS AND DURING 20 TRIALS IN S AND D CONTEXTS. ALSO SHOWN ARE MEANS OF PERCENTAGES OF ANTICIPATION AND RECALL OF THE HIGH MEANINGFULNESS RESPONSE MEMBERS AND LOW MEANINGFULNESS RESPONSE MEMBERS OF THE HL AND LH LISTS COMBINED AND OF THE HH AND LL LISTS UNDER RP OF 50–50

RP, List, and Int.	M/Freq.[a]	1–20	21–40	41–60	61–80	81–100	101–120	S	D
50–50									
HH									
2 sec.		27	42	46	47	47	47	49	40
24 hr.		21	40	43	44	44	44	45	19
HL									
2 sec.	High	28	52	58	55	61	59	57	47
and	Low	7	19	27	33	32	35	39	36
LH									
24 hr.	High	25	48	53	54	57	59	58	44
	Low	6	20	24	30	30	30	37	30
LL									
2 sec.		11	25	32	37	40	41	45	41
24 hr.		3	15	23	34	39	43	45	30
60–40									
HH									
2 sec.	More	26	53	54	57	57	57	56	28
	Less	20	26	29	34	34	34	39	20
24 hr.	More	29	53	54	58	58	58	61	52
	Less	15	23	27	31	31	31	34	36
HL									
2 sec.	More	29	67	79	78	78	77	72	76
	Less	6	12	13	16	16	17	26	23
24 hr.	More	22	58	66	71	70	72	68	45
	Less	9	13	17	18	22	22	24	23
LH									
2 sec.	More	7	19	28	30	29	32	40	28
	Less	20	43	52	57	61	60	57	50
24 hr.	More	9	28	35	34	45	48	52	38
	Less	18	34	41	48	44	43	40	38
LL									
2 sec.	More	6	29	47	52	55	60	63	58
	Less	2	7	12	19	22	25	32	35
24 hr.	More	7	26	37	44	56	56	52	38
	Less	3	9	12	16	20	22	28	24
70–30									
HH									
2 sec.	More	28	61	71	76	76	76	87	53
	Less	6	9	11	10	10	10	9	16
24 hr.	More	33	66	76	79	79	79	79	50
	Less	12	17	14	16	16	16	18	12
HL									
2 sec.	More	37	74	80	84	84	84	83	75
	Less	4	4	7	9	9	10	14	12
24 hr.	More	38	74	82	84	88	91	88	52
	Less	7	9	10	10	8	5	12	6
LH									
2 sec.	More	18	48	63	68	72	74	76	61
	Less	13	22	24	24	23	19	21	19
24 hr.	More	13	35	56	64	69	75	77	60
	Less	8	21	22	24	20	20	19	25

[a] M/Freq. designates meaningfulness of the alternative response members of the HL and LH lists combined under RP of 50–50, or the more frequent or less frequent response members under RP of 60–40 upward.

TABLE 2—*Continued*

RP	M/Freq.[a]	1–20	21–40	41–60	61–80	81–100	101–120	S	D
LL									
2 sec.	More	10	28	42	49	59	56	63	60
	Less	1	6	15	22	20	25	30	32
24 hr.	More	14	43	55	67	75	78	86	56
	Less	2	8	8	9	7	8	6	5
80–20									
HH									
2 sec.	More	45	76	87	90	90	90	94	90
	Less	6	7	6	4	4	4	4	6
24 hr.	More	52	76	85	86	86	86	86	60
	Less	4	10	9	10	10	10	12	15
HL									
2 sec.	More	41	81	90	94	96	98	97	66
	Less	1	3	2	2	1	0	2	7
24 hr.	More	52	84	91	95	96	97	97	82
	Less	1	2	3	3	1	1	2	7
LH									
2 sec.	More	16	61	81	86	90	91	90	67
	Less	5	7	6	6	5	6	7	11
24 hr.	More	21	50	68	75	79	83	91	55
	Less	9	17	14	14	12	10	6	9
LL									
2 sec.	More	19	48	70	80	86	91	96	73
	Less	2	2	2	2	1	2	2	1
24 hr.	More	27	67	82	89	91	92	94	56
	Less	1	4	2	2	2	3	1	2
90–10									
HH									
2 sec.	More	64	92	97	98	98	98	99	74
	Less	2	2	0	1	1	1	1	1
24 hr.	More	58	83	90	92	92	92	91	53
	Less	0	5	3	5	5	5	8	1
HL									
2 sec.	More	50	84	97	98	99	98	99	63
	Less	0	0	0	0	0	0	0	0
24 hr.	More	53	86	94	96	97	98	100	60
	Less	1	0	0	0	0	0	0	2
LH									
2 sec.	More	35	74	88	91	96	96	99	87
	Less	1	1	1	1	0	0	0	2
24 hr.	More	32	71	89	95	94	94	95	71
	Less	2	2	2	1	2	1	1	3
LL									
2 sec.	More	29	62	79	87	94	95	99	62
	Less	1	0	0	0	0	0	0	0
24 hr.	More	28	66	86	90	96	95	94	72
	Less	0	1	1	1	0	0	2	0

RESULTS

Percentages of anticipation or recall were determined for both the more frequent and less frequent response members of each unit. These were averaged across units for groups of *S*s; they were also examined for individual units for individual *S*s.

Across Units for Groups

Table 2 shows means of percentages of anticipation of the more frequent and of the less frequent response members of each list during successive blocks of 20 acquisition trials under RP of 60-40, 70-30, 80-20, and 90-10. The means for Trials 81–100 and 101–120 with the

HH list are estimates based on percentages of anticipation by *S*s during Trials 61–80. Also shown are means of percentages of recall of the response members during the trials in the S and D contexts both 2 sec. after and 24 hr. after the acquisition trials.

Under RP of 50-50, the HL and LH lists were equivalent. Therefore, means of percentages of anticipation and recall of the response members of high meaningfulness of these lists were averaged. Means of percentages of anticipation and recall of response members of low meaningfulness of these lists were also averaged. With the LL and HH lists, percentages of anticipation of alternative response members were averaged.

By the last 20 acquisition trials, except for the LL list, percentages of anticipation of the more frequent and less frequent response members generally summed to 90% or more. Therefore, analyses of variance were carried out only on percentages of anticipation and recall of the more frequent response members. One analysis was on percentages of anticipation during the acquisition trials. The other analysis was on percentages of anticipation during the last 20 acquisition trials and percentages of recall in the S and D contexts.

Acquisition.—In the analysis of variance on percentages of anticipation of the more frequent response members (Table 3), the variables are RP, meaningfulness of the more frequent response members (M), meaningfulness of the less frequent response members the same as or different from that of the more frequent response members (S-D), retention interval (I), and

TABLE 3

ANALYSES OF VARIANCE ON PERCENTAGES OF ANTICIPATION AND RECALL OF MORE FREQUENT RESPONSE MEMBERS DURING SIX BLOCKS OF ACQUISITION TRIALS, LAST 20 ACQUISITION TRIALS, AND TRIALS WITH STIMULUS MEMBERS ALONE IN SAME AND DIFFERENT CONTEXTS

Source	Acquisition Trials			Last 20 Trials, Trials in Same, Different Contexts		
	df	*MS*	*F*	*df*	*MS*	*F*
RP	4	145,963.94	160.62**	4	72,408.14	105.10**
Meaningfulness (M)	1	138,125.64	151.99**	1	9,058.96	13.15**
M Same or Different (S-D)	1	4,341.03	4.78*	1	2,943.50	4.27*
Retention Interval (I)	1	178.24		1	902.88	1.31
RP × M	4	2,664.17	2.93*	4	1,663.50	2.41*
M × S-D	1	7,966.62	8.77**	1	8,802.76	12.78**
RP × M × S-D	4	2,636.93	2.90*	4	2,053.17	2.98*
RP × M × S-D × I				4	2,382.57	3.46**
Error (b)	280	908.78		280	688.94	
Trials or Test Conditions (T)	5	102,468.08	1,124.19**	2	39,800.64	103.09**
T × RP	20	1,053.11	11.55**	8	1,149.64	2.98**
T × M	5	3,027.53	33.22**			
T × S-D	5	213.19	2.34*			
T × I				2	1,981.56	5.13**
T × RP × M	20	400.58	4.40**			
T × M × S-D	5	899.55	9.87**			
T × RP × M × S-D	20	164.80	1.81*			
Error (w)	1240[a]	91.15		560	386.06	

Note.—Interaction *F*s not significant at .05 are omitted.
[a] $df = 1{,}400 - 160$ estimated values.
* $p \leq .05$.
** $p \leq .01$.

(blocks of) trials (T). M and S-D subsume the HH, HL, LH, and LL lists. None of the *F*s involving I was significant. Thus, the different groups of *S*s at the two retention intervals were equivalent and, in descriptions of effects of RP, M, S-D, and T, I can be disregarded.

The significant *F* for RP (Table 3) reflected an overall direct relationship between percentages of anticipation of the more frequent response members and actual percentages of their occurrence. The relationship holds at each of the blocks of trials, for each of the lists separately, and for each of the lists separately at each block of trials.

The significant RP × M × S-D and T × RP × M × S-D interactions are of primary importance. The former interaction reflects differences among RP in orders of means of the lists and in absolute sizes of differences between means of adjacent lists. Under RP of 60-40, the order is HL > HH > LL > LH with approximately equal differences between adjacent lists. Under RP of 70-30, 80-20, and 90-10, respectively, the orders are HL > HH > LH > LL, HL > HH > LL ≑ LH, and HH ≑ HL > LH ≑ LL. Negligible differences between adjacent lists are indicated by ≑, with the list with the larger mean to the left. In addition to differences among RP in order, as RP change from 60-40 upward differences between the HL and HH lists and between the LL and LH lists decrease. Differences between the former two and the latter two lists also decrease but less markedly.

In the T × RP × M × S-D interaction, inversions from the order across trials at one or more of the six blocks of trials occur under each of the RP. The inversions are typically in the orders of the HL and HH lists and of the LL and LH lists, rather than in the order of HL and HH relative to LL and LH.

The patterns of these interactions suggest that differences due to M occur at each level of S-D under RP from 60-40 upward. Differences due to S-D occur at each level of M under RP of 60-40 and, perhaps, under RP of 70-30. Differences due to S-D also occur with more frequent response members of high meaningfulness under RP of 80-20 but do not occur under any other combination of RP and M. The patterns of the other interactions, both significant and nonsignificant, are consistent with this description of effects of M and S-D as RP change from 60-40 upward as is the greater overall effect of M than of S-D.

Under RP of 50-50, percentages of anticipation of response members of high meaningfulness of the HL, LH lists increase most rapidly and those of low meaningfulness increase least rapidly. Percentages of anticipation of response members of the HH and LL lists increase at intermediate rates with HH > LL.

Last 20 acquisition trials, trials in the S and D contexts.—In the analysis of variance on percentages of anticipation during the last 20 acquisition trials and of recall during trials in the S and D contexts (Table 3), trials are replaced by test conditions (also T). RP is again significant and again reflects an overall direct relationship between percentages of anticipation or recall of the more frequent response members and actual percentages of their occurrence.

As a general indication of maintenance and generalization, the significant T × RP interaction is important. Under each of the RP the percentage of anticipation during the last 20 trials and the percentage of recall during the trials in the S context were essentially equal. However, during trials in the D context, progressively larger decrements in percentages of recall occur under RP from 50-50 upward.

The significant RP × M × S-D interaction is more important. Under RP of 60-40, the order is $\mathrm{HL} > \mathrm{LL} \doteq \mathrm{HH} > \mathrm{LH}$; under RP of 70-30 and 80-20, the order is $\mathrm{HL} > \mathrm{HH} \doteq \mathrm{LL} \doteq \mathrm{LH}$; and under RP of 90-10, the order is $\mathrm{LH} \doteq \mathrm{HL} \doteq \mathrm{LL} \doteq \mathrm{HH}$. Differences due to M occur under RP from 60-40 through 80-20 but only with response members of different meaningfulness. Differences due to S-D occur under RP of 60-40 at each level of M, and under RP of 70-30 with more frequent response members of low meaningfulness. The patterns of lower order interactions involving RP, M, and S-D and the greater overall effect of M than of S-D are consistent with this description of effects of M and S-D as RP change from 60-40 upward.

The nonsignificant T × RP × M × S-D interaction suggests that the order under each RP across T holds during the last 20 acquisition trials, trials in the S context, and trials in the D context separately. In the significant RP × M × S-D × I interaction, under each of the RP the order of lists at one retention interval was reasonably congruent with the order across intervals, but the order at the other retention interval was different. The congruent orders were not consistently at one interval or the other.

During the last 20 acquisition trials under RP of 90-10 and 80-20, anticipations with all four lists exceed 90% and 80%, respectively. Under RP of 70-30, anticipations with all lists but the LL list exceed 70%. Under RP of 60-40, anticipations of the more frequent response members of the HL list are at about 75%. Anticipations with the HH and LL lists are at or slightly below 60%, and anticipations of the more frequent response members of the LH list are at about 40%.

Under RP of 50-50, anticipations of the more meaningful and less meaningful response members of the HL, LH lists are seemingly stable at about 59% and 32%, respectively. Anticipations of response members of the HH list are seemingly stable at about 46%. However, anticipations of response members of the LL list, which are at about 42%, may have been still rising.

During trials in the S context, recall of the more frequent response members of all lists is above 90%, 80%, and 70% for acquisition under RP of 90-10, 80-20, and 70-30, respectively. For acquisition under RP of 60-40 and 50-50, percentages of recall with each list are essentially the same as percentages of anticipation during the last 20 acquisition trials.

During the trials in the D context, for all lists under RP from 60-40 upward and for the HH and LL lists under RP of 50-50, percentages of recall of the more frequent response members are below their actual percentages of occurrence during acquisition trials. Under RP of 50-50, recall of response members of high meaningfulness of the HL, LH lists is slightly below 50%; recall of the response members of low meaningfulness does not change.

Individual Units for Individuals

Although not shown here, numbers of *S*s for whom hierarchies were obtained for three or four units were determined for each of the combinations of RP, lists, and I during the last 20 trials and trials in the S and D contexts. The criterion of presence of an hierarchy was that employed by Goss and Cobb (1966, p. 226).

During the last 20 trials and trials

in the S and D contexts, the numbers of *S*s who exhibit hierarchies for three or four units decrease from around 50% under RP of 50-50 to less than 10% under RP of 90-10.

In separate Sutcliffe χ^2 analyses of differences between and among numbers of *S*s who exhibit hierarchies during the last 20 trials and during trials in the S and D contexts, RP was the only variable significant in more than one of the three analyses. Thus, neither lists as represented by M and S-D nor I seemingly influence numbers of *S*s who exhibit hierarchies either across RP or under one or more RP. At least under RP of 50-50, 60-40 and 70-30, more *S*s exhibit hierarchies during the trials in the S and D contexts than during the last 20 trials.

As previously (Goss & Cobb, p. 227), stability of hierarchies was assessed by phi coefficients for presence or absence of hierarchies in pairwise correlations involving the last 20 trials and trials in the S and D contexts. More coefficients significant at .05 occur for correlations involving the S and D contexts than for correlations involving the last 20 trials and the S or the D context. Systematic differences among lists are not apparent under any of the RP.

DISCUSSION

Formation of response hierarchies across units for groups of *S*s during the acquisition trials is examined and then maintenance, generalization, and retention of those hierarchies. For individual units for individual *S*s, lists had little or no influence on these phenomena. Accordingly the results are not discussed except to note that, in general, those for the HH list are consistent with Goss and Cobb's (1966) results.

Formation.—On the basis of prior findings of effects of RP with an HH list and of meaningfulness of response members of PA units in a 1:1 pattern of pairings (Goss & Nodine, 1965), under RP of 60-40 upward the expectation was of an HL > HH > LL > LH order of rates of acquisition. However, the significant RP × M × S-D interaction suggests differences among RP in orders of the lists in a pattern of change from the expected order under RP of 60-40 to an HH ≐ HL > LH ≐ LL order under RP of 90-10.

More specifically, under all RP from 60–40 upward, M seemingly has a direct effect on rate of acquisition at both levels of S-D. S-D has direct effects at both levels of M under 60-40, but only with response members of high meaningfulness under RP of 70-30 and 80-20. Effects of S-D disappear under RP of 90-10. These restrictions on effects of S-D underlie the failure to obtain the expected order of rates of acquisition of the lists under each of the RP.

Under RP of 50-50, as expected, response members of high meaningfulness were acquired more rapidly than those of low meaningfulness in comparisons both with the HL, LH lists and with the HH and LL lists.

During the last 20 acquisition trials, the obtained order of lists across RP is HL > HH ≐ LL ≐ LH. The changes in orders under RP from 60-40 upward suggest that M has a direct effect on percentages at which anticipations stabilize only under RP of 60-40, 70-30, and 80-20 with lists whose response members are of different meaningfulness. That anticipations of more frequent response members of HH and LL lists stabilize at the same levels is consistent with findings for PA lists with 1:1 pairings of stimulus members and response members (Goss & Nodine, 1965).

Previous findings (Goss, 1965; Goss & Cobb, 1966; Goss & Sugerman, 1961) of anticipations of more frequent response members of the HH list above 90%, 80%, and 70% during the last 20 trials under RP of 90-10, 80-20, and 70-30, respectively, are replicated. In addition, with the LL list under RP of 70-30 as a possible exception, these relationships of percentages of anticipation and of actual occurrence apparently hold with HL, LL, and LH lists. With the HH list under RP of 60-40 and 50-50, as previously, anticipations were at about 60% and slightly below 50%, respectively. The same relationships hold with the LL list.

With the HL and LH lists under 60-40, anticipations of the more frequent response members were above and below 60%, respectively. With the HL, LH lists under RP of 50-50, anticipations of response members of high and of low meaningfulness were seemingly stable above 50% and below 50%, respectively.

Maintenance.—Under the same RP, the order of means of percentages of recall of the more frequent response members during the trials in the S context is generally the order of the last 20 acquisition trials. Thus, with the HH list, as previously (Goss & Cobb, 1966), and also with HL, LL, and LH lists, hierarchies hold through 20 trials in the S context.

Generalization.—With the change from the S context to the D context, progressively larger decrements in percentages of recall of the more frequent response members occur under RP from 50-50 upward. Also, decrements at the 24-hr. retention interval are greater than those at the 2-sec. interval.

None of the interactions involving test conditions was significant to suggest that, in general, decrements from the S context to the D context were about equal under each of the RP both with the HH list and with the HL, LL, and LH lists. Retention interval disregarded, under each of the RP the orders of the means of the lists are essentially the same during trials in the D context as during the last 20 trials and trials in the S context.

In Goss and Cobb's first experiment, the decrements obtained with the HH list from trials in the S context to those in the D context were generally smaller than the decrements obtained with the HH list in the present experiment. In Goss and Cobb's second experiment, no decrements occurred. Very likely, the reason for the larger decrement in the present experiment both with the HH list and with the HL, LL, and LH lists is the use of the stimuli alone on a single sheet of paper in contrast to the previous use of booklets with one stimulus on each page. With the sheet of paper, *S*s can see preceding and succeeding stimuli and preceding responses. Thus, it represents a greater change from the conditions of trials in the S context than did the booklets.

Retention.—Whether trials in the S and D contexts were administered immediately after or 24 hr. after the acquisition trials made little difference. In Goss and Cobb's experiments, also, retention interval had no or negligible effects.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Goss, A. E. Paired-associates learning with varying relative percentages of occurrence of alternative response members: Influence of instructions. *J. exp. Psychol.*, 1965, **70**, 51–56.

Goss, A. E., & Cobb, N. J. Formation, maintenance, generalization, and retention of response hierarchies. *J. exp. Psychol.*, 1966, **71**, 218–231.

Goss, A. E., & Nodine, C. F. *Paired-associates learning: The role of meaningfulness, similarity, and familiarization.* New York: Academic Press, 1965.

Goss, A. E., & Sugerman, M. E. Paired-associates learning with varying relative percentages of occurrence of alternative response members. *J. exp. Psychol.*, 1961, **62**, 24–34.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

(Received July 1, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 282–288

EFFECTS OF NONRANDOM INTERMITTENT REINFORCEMENT SCHEDULES IN HUMAN-EYELID CONDITIONING[1]

WILLIAM F. PROKASY
University of Utah

RONALD A. CARLTON
Washington & Jefferson College

AND J. DAVID HIGGINS
Pennsylvania State University

Random, single alternation, double alternation, and 2 nonrandom (1 biased toward runs and the other biased toward alternation) reinforcement schedules were employed in human classical eyelid conditioning. The major findings were that with the same alternation pattern some Ss responded according to the pattern and others did not and that the random and nonrandom probabilistic schedules yielded data generally consistent with a linear operator model, but that parameters depended upon experience with the specific schedule. It was concluded that the failure of parameters to remain invariant over early training necessitates more extensive training to achieve stability and that more attention will have to be paid to individual-S response characteristics, since the same reinforcement schedules did result in lawfully different behavior in different Ss.

A characteristic common to all "strength" models of learning (e.g., Bush & Mosteller, 1951; Hull, 1943; Spence, 1956) is that the tendency to make a response increases following a reinforcement and decreases following a nonreinforcement. Mean response strength resulting from an intermittent reinforcement schedule then represents the averaged effects of increments and decrements in response strengths, with terminal performance reflecting the reinforcement ratio, the incremental effect of a reinforcement (θ), the decremental effect of a nonreinforcement (θ'), and the limits of performance, A and A′, resulting from, respectively, an indefinitely long sequence of reinforcements or nonreinforcements.

Relatively little is known about the extent to which the incremental-decremental model (hereafter, model) provides an adequate description of the effects of intermittent reinforcement in human conditioning, although most of the available data are consistent with it. For example, as the ratio of reinforcements to nonreinforcements decreases, response probability in human-eyelid conditioning decreases (e.g., Grant & Schipper, 1952; Humphreys, 1939; Kimble, Mann, & Dufort, 1955; Reynolds, 1958). In addition, when patterns of reinforcement and nonreinforcement, such as single or double alternation, are employed in eyelid conditioning with standard instructions, *S*s do not appear to respond to the patterning per se (Grant, Riopelle, & Hake, 1950; Hartman & Grant, 1962; Hickok & Grant, 1964). On the other hand, when *S*s are given specific instructions which orient them to the re-

---

[1] This research was supported by funds from National Science Foundation Grant GB-3875 to the senior author. We would like to acknowledge the help provided by Martha Harsanyi, a participant in the Undergraduate Research Participation program at the Pennsylvania State University under Grant GY-975 from the National Science Foundation.

inforcement schedule, they do respond more frequently on reinforced than on nonreinforced trials, a clear indication of responding to the pattern (Hartman & Grant, 1962; Hickok & Grant, 1964). Similar response patterning has been observed with the GSR even in the absence of specific instructions to orient to the reinforcement pattern (Longenecker, Krauskopf, & Bitterman, 1952).

The purpose of this paper is to examine the effects of five 50% reinforcement schedules in order to help determine how much generality the model has in human-eyelid conditioning under standard instructional circumstances. The experiment to be reported does not constitute a "test" of the model; rather it is designed to provide data which have relevance to its appropriateness as a model for human conditioning.

METHOD

Apparatus.—The apparatus for picking up and recording eyelid responses is described elsewhere (Prokasy & Whaley, 1961). A modification was that response signals were amplified and recorded with an Offner Type RC dynograph operating at a paper speed of 250 mm/sec. Additional changes were that *S*s were seated in a 6 × 6 ft. cubicle, well lighted by a pair of overhead fluorescent lamps, and that two *S*s were run simultaneously in different cubicles with the same reinforcement schedule.

The CS was a 500-msec. increase in brightness of a .75-in. diameter standard translucent red cut-glass jewel. It was illuminated from behind by a 7.5-w. bulb. The CS source was mounted on a 1 × 2 ft. panel attached to the cubicle wall at approximately 60 in. in front of *S* at eye level. The UCS was a 50-msec. puff of nitrogen delivered to *S*'s right cornea 500 msec. after CS onset. It was delivered through a 4.5-mm. diameter jet located approximately .50 in. from *S*'s eyelash and was sufficient to support a 150-mm. column of mercury.

Interstimulus interval, CS, and UCS durations were controlled by Tektronix timers. Equipment cues were masked by a combination of white noise and a tape recording of the dynograph paper drive, the combination being 77 db. (re .0002 dyne/cm^2).

Subjects.—The *S*s were 165 volunteers from introductory psychology courses at the Pennsylvania State University who participated in the study to fulfill part of some optional course credits. Of these, four were lost through apparatus failure of *E* error.

Experimental design.—The independent variable was the type of 50% intermittent reinforcement schedule. The *S*s were assigned unsystematically to one of five 300-trial schedules. Group R, $N = 36$, received a random schedule within the restraint that each 150 trials had exactly 75 reinforced and 75 nonreinforced trials. Group SA, $N = 29$, received a single alternation schedule. Group DA, $N = 34$, received a double alternation schedule. Group 8R, $N = 32$, received a schedule in which the likelihood of a reinforcement following a reinforcement was .8 and the likelihood of a nonreinforcement following a nonreinforcement was .8. This schedule tended to produce runs of reinforced or nonreinforced trials. Group 2R, $N = 30$, received a schedule in which the likelihood of a reinforcement following a reinforcement or a nonreinforcement following a nonreinforcement, was .2. This schedule tended toward alternation between reinforcements and nonreinforcements. Groups 2R and 8R schedules were restricted to 75 reinforced and nonreinforced trials in each 150 trial block.

After being placed in the cubicle, all *S*s received a set of instructions, by tape recorder, in which they were instructed to sit quietly, not to try to figure out the experiment, and to let their reactions take care of themselves. They then received a random order of 12 adaptation trials, 6 CS alone and 6 UCS alone, followed by 300 conditioning trials. The average intertrial interval was 10 sec., with values ranging from 6 to 14 sec.

A CR was defined as a 1.0-mm., or greater, deflection of the recording pen between 148 and 524 msec. after CS onset. A 1.0-mm. pen deflection corresponded to an eyelid movement of approximately .9 mm.

RESULTS

Group performances stabilized by the 150th trial. Table 1 provides response probabilities and *SD*s for the five groups across Trials 151–300. Groups DA and 2R had terminal performance levels above $p = .6$, while

TABLE 1

MEAN RESPONSE PROBABILITIES AND *SD*s FROM TRIAL 151 TO 300

	Group				
	R	2R	8R	SA	DA
Mean	.454	.600	.418	.478	.615
Variance	.125	.120	.064	.068	.069

Groups SA, 8R, and R had terminal probabilities below $p = .5$. However, overall differences among the means were not significant, F (4, 156) = 1.99.

Groups DA and SA.—With single and double alternation schedules the first question is whether or not there is evidence that *S*s responded to the pattern of reinforcements rather than simply to the immediate aftereffects of reinforcement and nonreinforcement. Table 2 reports mean performance in three blocks of 100 trials each for Group SA and in two blocks of 150 trials each for Group DA. In both groups probability on reinforced trials was greater than that on nonreinforced trials, F (1, 28) = 15.17 for Group SA and F (1, 33) = 5.53 for Group DA. Since the incremental-decremental model would predict the reverse ordering, higher probability on nonreinforced trials, there is evidence of responding to the pattern per se.

TABLE 2

MEAN RESPONSE PROBABILITIES ON NONREINFORCED AND REINFORCED TRIALS

Trials	Nonreinforced		Reinforced	
	Mean	Variance	Mean	Variance
Group SA				
1–100	.228	.031	.420	.011
101–200	.382	.062	.551	.094
201–300	.399	.076	.558	.087
Group DA				
1–150	.460	.086	.516	.081
151–300	.563	.095	.666	.068

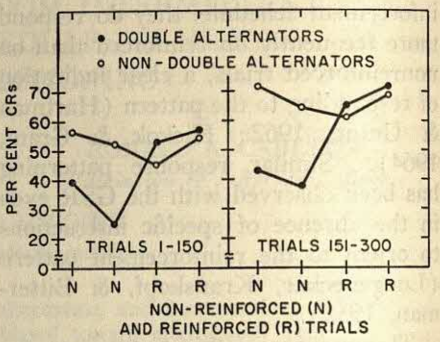


FIG. 1. Percentage of CRs across the four-trial double alternation sequence for two subsets of *S*s from Group DA.

Some *S*s in Group DA tended to double alternate. The *S* was defined as a double alternator if, averaged over the *last* 150 trials, his response probability to both the first and second nonreinforced trials in the four-trial sequence was less than that to the first and second reinforced trials. Fourteen of the 34 *S*s met this criterion, and Fig. 1 reports response probability for both subsets of *S*s in two 150-trial blocks. Response probabilities over the four-trial sequence differed for both the double alternators and the nondouble alternators, F (3, 39) = 17.75 and F (3, 57) = 16.00, respectively.

Several features of Fig. 1 should be noted. First, the 20 who did not double alternate behaved as the model would predict: decrements following each nonreinforcement and increments following each reinforcement. Second, the two subsets, though separated by a criterion based on performance in the last 150 trials, yield similar functions in the first 150 trials: this not only provides an index of the reliability of the separation, but also suggests that the dominant response style is determined early and maintained throughout training. Third, double alternators tended to increase response probability following the first reinforcement and decrease it following the first nonreinforcement. Only the decrement following the first nonreinforcement is reliable, however, with 13 and 11, respectively, of the 14 double alternators showing a decrement in the first and last blocks of 150 trials, χ^2 (1) = 10.28 and 4.57, respectively. In the first and last blocks only eight and nine, respectively, showed increments following the first reinforcement. Finally it is to be noted that the difference between the two subsets of *S*s resides in the different levels of responding on the nonreinforced trials: response probability on reinforced trials was approximately the same for both subgroups.

Further evidence of differences between the subsets of double alternators and nondouble alternators in Group DA resides in response latency. For the double alternators latencies were longer on the nonreinforced than on the reinforced trials, the values being 385, 387, 372, and 372 msec., *F* (3, 39) = 3.19. For the nondouble alternators the mean latencies of 375, 371, 373, and 368 msec. were approximately the same as the latencies on the reinforced trials for the double alternators and did not differ among themselves, *F* (3, 57) = .65. Comparable latency effects were observed in Group SA. Of 22 *S*s who yielded a higher response probability on reinforced than on nonreinforced trials, 16 had a shorter latency on reinforced trials, $\chi^2(1) = 4.56$, $p < .05$. No reliable effects were observed on the 7 *S*s of Group SA who did not show alternation behavior.

To provide a basis for comparison with past research, response probabilities on reinforced and nonreinforced trials in Group DA were calculated across Trials 33–60. Although probability was higher on reinforced trials, the means of .457 and .395 did not differ reliably, *F* (1, 33) = 1.82, consistent with past research (Grant et al., 1950; Hartman & Grant, 1962; Hickok & Grant, 1964). Having divided *S*s into subsets of double and nondouble alternators, it was possible to obtain a different picture of this result. Of the 14 double alternators, 13 yielded a higher probability on reinforced trials over Trials 33–60, $\chi^2(1) = 10.28$, while of 17 nondouble alternators (3 ties discounted) 14 had a higher probability on nonreinforced trials, χ^2 (1) = 7.12. Thus, the failure to obtain overall differences on reinforced and nonreinforced trials can be attributed to the averaging of probabilities across subsets of double and nondouble alternators.

Groups 8R, 2R, and R.—Since all three groups received a probabilistic

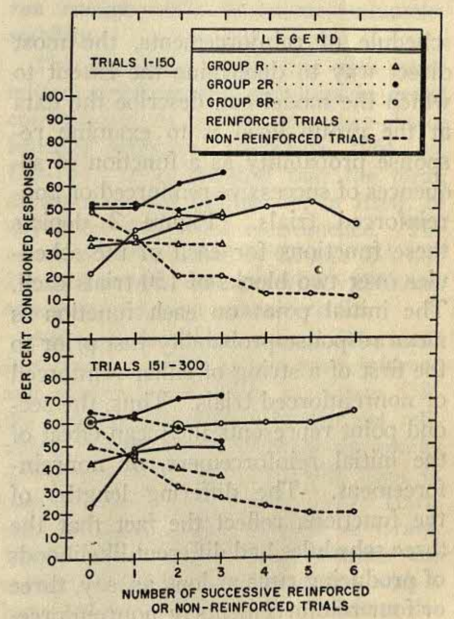


FIG. 2. Response probability as a function of the number of preceding successive reinforced or nonreinforced trials in Groups R, 2R, and 8R. (Number "0" represents the trials on which the first of a succession of either reinforced or nonreinforced trials occurred.)

TABLE 3

ANALYSIS OF VARIANCE OF SEQUENCES OF REINFORCEMENT AND NONREINFORCEMENT

Source	Group						Error Term
	R		8R		2R		
	df	*F*	*df*	*F*	*df*	*F*	
Trials (T)	1	13.32*	1	14.52*	1	11.94*	a
CS+ vs. CS⁻ (C)	1	25.23*	1	63.07*	1	19.77*	b
Sequence (S)	3	1.04	6	4.69*	3	2.31	c
T × C	1	.13	1	.03	1	.66	d
T × S	3	9.69*	6	1.51	3	3.33*	e
C × S	3	18.77*	6	53.99*	3	9.14*	f
T × C × S	3	.63	6	1.62	3	1.22	g
Error Terms[a]							
I × T[b] (a)	35	(.0584)	31	(.1612)	29	(.0700)	
I × C (b)	35	(.0088)	31	(.1529)	29	(.0247)	
I × S (c)	105	(.0051)	186	(.0165)	87	(.0166)	
I × T × C (d)	35	(.0095)	31	(.0384)	29	(.0152)	
I × T × S (e)	105	(.0055)	186	(.0158)	87	(.0220)	
I × C × S (f)	105	(.0119)	186	(.0400)	87	(.0264)	
I × T × C × S (g)	105	(.0105)	186	(.0211)	87	(.0161)	

[a] MS_e in parentheses.
[b] I designates individual *S*s.
* $p < .05$.

schedule of reinforcements, the most direct way to determine the extent to which the model can describe the data at the group level is to examine response probability as a function of sequences of successive reinforced or nonreinforced trials. Figure 2 depicts these functions for each of the schedules over two blocks of 150 trials each. The initial point on each function is mean response probability just prior to the first of a string of either reinforced or nonreinforced trials. Thus, the second point represents the mean effect of the initial reinforcement, or nonreinforcement. The differing lengths of the functions reflect the fact that the three schedules had different likelihoods of producing runs as long as, say, three or four reinforcements or nonreinforcements in a row. An analysis of variance was conducted separately on the data of each group, and is summarized in Table 3.

The same general trends emerged from all sets of data. As anticipated, the means increased from the first to the second trial block, and mean response probability was less following a nonreinforced trial than following a reinforced trial. The sequence effect, a decrease in probability across combined reinforced and nonreinforced sequences, was significant only in Group 8R. The significant Trials × Sequence effect in Groups R and 2R[2] is attributable to the mean rise across both positive and negative sequences in early training as contrasted with no mean change across the same sequences late in training. Successive reinforcements, in general, led to increases in performance while successive nonreinforcements, in general, resulted in performance decreases, thus accounting for the significant Reinforcement × Sequence interaction. This last effect means that at this molar level of analysis, the model can describe the data of all three groups.

---

[2] Following the suggestion of Greenhouse and Geisser (1959), it may be advisable not to interpret the T × S *F* ratio in Group 2R as significant. To allow for possible unequal covariance matrices in a repeated-measures analysis, they suggest the conservative adjustment of reducing the numerator *df* to one, which would not permit rejection of the null hypothesis at the .05 level with an *F* of 3.33.

DISCUSSION

The major purpose of this experiment was to determine to what extent the incremental-decremental model provides a general description of the sequential behavior of human *S*s in a representative eyelid conditioning situation. While the gross sequential data of Groups 2S, 8S, and R are consistent with the model, it is evident that the pattern effects observed in Groups SA and DA necessitate, at the very least, additional assumptions. Thus, there do exist intermittent reinforcement schedules which result in conditioning performance not compatible with the model, even when *S*s are not given explicit instructions concerning the schedule.

Although individual differences and complexities of parameter estimation for specific models preclude more extensive analyses of the data at this time, the sequential curves of Groups 8R, 2R, and R in Fig. 2 do have further implications if the model is to be applied. To begin with, the three schedules result in mean functions which require different parameters. Group 8R has a larger θ and θ' than does Group R, as evidenced by the fact that the sequential curves of the former group cross those of the latter. The higher mean response level of Group 2R implies that A, and perhaps A′, are higher for that group than for the other two. Even though the mean levels are higher, it is to be noted that the sequential curves of Group 2R parallel those of Group R, suggesting that θ is larger, and θ' smaller, in the former.

In itself the difference in parameters, assuming the model to work, is no hardship. It does have two related consequences, however, which merit elaboration. First, it has been shown that *S*s do not enter the situation with a set of fixed parameters. In particular, while it may be possible to account for *S*'s behavior with an incremental-decremental model, the parameter values are determined, not by the general situation, but by information obtained by *S* from the reinforcement schedule itself. Thus, even if performance appears to be subject to the immediate effects of reinforcement and nonreinforcement, the parameter values which result in that outcome are determined by experience with the specific reinforcement sequence. The related consequence is that if parameter values are a function of the reinforcement sequence, then some minimum number of conditioning trials must be presented before they can be determined. Coupled with individual differences, variable parameters during early training strongly suggest that the data of early training in human conditioning have little to offer for the analysis of performance with an intermittent reinforcement schedule. The practical implication for *E*s is evident: run enough trials to assure parameter stability.

This study also shows that with intermittent reinforcement schedules in eyelid conditioning, as well as in a two-choice learning situation (Anderson, 1964, pp. 129–144), it is doubtful that the mean asymptotic level reached is, in itself, of any great theoretical consequence. The common level reached by Groups DA and 2R, for example, obscures wholly different sequential response characteristics.

Although there were overall differences in response probability between reinforced and nonreinforced trials in Group DA, the fact that the difference was not reliable across Trials 33–60 keeps the results of the present study in line with those of past studies (Grant et al., 1950; Hartman & Grant, 1962; Hickok & Grant, 1964). The present study extends the earlier results by illustrating that with a large number of trials it is possible to define subsets of double alternators and nondouble alternators and that the earlier failure to obtain differences between reinforced and nonreinforced-trial probabilities may simply have resulted from an averaging across two such subsets of *S*s.

Particularly noteworthy, in any event,

is the fact that response probability on reinforced trials in Group DA was not at all contingent on whether or not S was classed as a double alternator: the pattern effect seems exclusively to be an inhibition of responding on the nonreinforced trial. Further evidence of this is provided by the fact that response latencies were reliably longer on nonreinforced than on reinforced trials for the double alternators. Similar effects in eyelid conditioning were obtained by Hickok and Grant (1964), but not by Hartman and Grant (1962), under instructions to orient to the reinforcement pattern. It seems probable that response inhibition on nonreinforced trials arises from a self-instructed orientation to the reinforcement pattern. In this regard, only 10 of 34 Ss failed to report detecting the double alternation reinforcement pattern, of which only two were classed as double alternators.

The data of Groups SA and DA provide complications in yet another way for model testing. In both groups some Ss performed according to expectation from the model, and some according to expectation from the reinforcement pattern. Whether or not verbal mediation can account for these disparities within the same experimental setting, the important point is that homogeneous treatment provides no assurance that all Ss will perform according to the same laws. It is conceivable that such between-Ss differences arise from different self-instructions and that more constraints added to the experimental situation will eliminate them. Alternatively, it may be that, in addition to obtaining more extensive data, a proper data analysis can be made only on an individual-S basis.

REFERENCES

Anderson, N. H. An evaluation of stimulus sampling theory: comments on Professor Estes' paper. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964. Pp. 129–144.

Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.,* 1951, **58,** 413–423.

Grant, D. A., Riopelle, A. J., & Hake, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.,* 1950, **40,** 53–60.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.,* 1952, **43,** 313–320.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometrika,* 1959, **24,** 95–112.

Hartman, T. F., & Grant, D. A. Effects of pattern of reinforcement and verbal information on acquisition, extinction, and spontaneous recovery of the eyelid CR. *J. exp. Psychol.,* 1962, **63,** 217–226.

Hickok, C. W., & Grant, D. A. Effects of pattern of reinforcement and verbal information on acquisition and extinction of the eyelid CR. *J. gen. Psychol.,* 1964, **71,** 279–289.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.,* 1939, **25,** 141–158.

Kimble, G. A., Mann, L. I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.,* 1955, **49,** 407–417.

Longenecker, E. G., Krauskopf, J., & Bitterman, M. E. Extinction following alternating and random reinforcement. *Amer. J. Psychol.,* 1952, **65,** 580–587.

Prokasy, W. F., & Whaley, F. L. The intertrial interval in classical conditioning. *J. exp. Psychol.,* 1961, **62,** 552–559.

Reynolds, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.,* 1958, **55,** 335–341.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received July 2, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 289–293

REPETITION AND THE PREPERCEPTUAL INFORMATION STORE

MICHAEL T. TURVEY

Ohio State University

The Sperling (1960) procedure of partial report was used with tachistoscopically presented digit slides of 50-msec. duration. 1 slide was repeated 54 times with a digit slide of a nonrepeat series interpolated between repetitions. No cumulative effect in terms of percentage of recall was obtained. A 2nd experiment using interpolated letter slides replicated this observation. An interpretation in terms of the lack of a structural change with repetition of a nonencoded stimulus was offered.

As demonstrated by Hebb (1961), Hellyer (1962), and Melton (1963), the memory trace in the short-term memory (STM) paradigm of the Peterson and Peterson (1959) model is considerably affected by the frequency of repetition and is, therefore, in this case, similar to the memory trace in the long-term memory (LTM) paradigm. To account for this fact Hebb (1961) postulated that a stimulus induces not only a dynamic trace but simultaneously a structural trace which can be cumulative, a statement well respected by Melton (1963). However there is some question as to what are the necessary conditions for a given stimulus item to induce a structural trace or, in another vernacular, to be placed into some form of relatively permanent storage. Broadbent (1963) speaks of the necessity for a stimulus to be encoded in order that it might be stored. With respect to a visual display of letters or digits encoding might be defined as the translation of that display into a verbal form for future retrieval. The nature of the Sperling (1960) paradigm is such that *S* on presentation of the display does not know which aspect of the display is to be retrieved and this together with the limited duration of the displays allows for the assumption that the Sperling procedure deals with the retrieval of a nonencoded stimulus. This contrasts with the retrieval of an encoded stimulus in the more usual STM protocol of the Peterson variety. If storage is predicated upon encoding then repetition of a stimulus display in the Sperling paradigm should not produce the same cumulative effect as evidenced with the Peterson paradigm.

EXPERIMENT I

Method

Materials.—Forty slides were constructed which consisted of three rows of five digits. The digits were randomized so that no two rows were the same.

A Research Instrument Laboratories Model V-1459 projecting tachistoscope equipped with an Atcotronic timer was used to project the stimulus slides onto the viewing screen. The presented field subtended a visual angle of 3.75° vertically × 5.25° horizontally at the viewing distance of 8½ ft. One tachistoscope channel projected a dark preexposure field constantly onto the screen to reduce the detrimental effects of brightness contrast. The center of the preexposure field was marked by a faint but discernable black X, the stimulus display being superimposed upon this field. Initiation of the stimulus display and the poststimulus cue sequence was by means of an *S*-controlled switch. The display terminated after a 50-sec. duration and was followed by one of three possible cues. These were tones generated by a Jackson

Model, 655 audio-oscillator and were transmitted to *S* by means of earphones. A white noise generator (Grason-Stadler Model 901B) maintained a constant background throughout the experimental procedure.

A Tektronix Model 160-A power supply provided the energy for three Tektronix Model 162 waveform generators, which together with associated pulse generators were set for given delays of the poststimulus cue. The three delay intervals of 0, 300, and 700 msec. could be interchanged by a 3-point switch controlled by *E*. A Hunter timer was used to maintain a continuous monitoring of the delay interval.

Subjects.—The *S*s were seven males and one female aged between 18 and 22 yr. engaged in the introductory course in general psychology at Ohio State University.

Procedure.—Each *S* received 3 days of preexperimental training. The first day involved establishing the auditory poststimulus cues and orienting *S* to the procedure. The cues were designed as high, middle, and low and the tones were adjusted individually for each *S* to obtain maximum possible ease of differentiation. The *S* was told that the study was testing his ability to retain a visual image. The stimulus format (i.e., a slide, with 3 rows of 5 digits) was presented onto a screen for *S*'s perusal, *S* being informed that no digit would appear twice in the same row. The procedure for producing a stimulus display involved *S* pressing a key following *E*'s statement of "ready." This allowed for *S* to be prepared for the visual presentation. The poststimulus cue demanded that *S* report one of the three rows in the fashion of Sperling's (1960) "partial report" technique. The interstimulus interval was never less than 10 sec. and never greater than 15 sec.

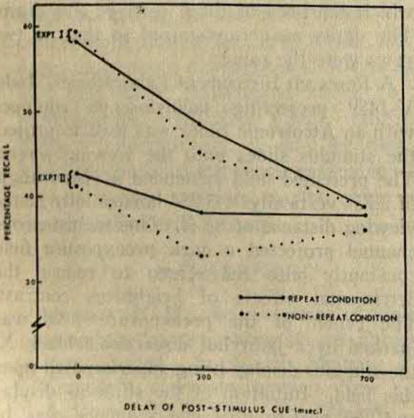


FIG. 1. Recall with respect to delay of poststimulus cue for Exp. I and Exp. II. (Each data point represents 144 observations.)

On the second and third days *S* received, respectively, 54 and 72 slide presentations. The rows to be recalled and the poststimulus delays were randomized. Following each presentation *S* recorded his recall on a response sheet which contained five blocks for each presentation. The *S* was required to record the digits in the sequence that they had appeared in on the slide and following each recording *S* pushed his response sheet under a cover sheet so that previous recordings were no longer visible to him. While *S* recorded his recall *E* changed slide, delay interval and the poststimulus cueing tone where necessary.

The purpose of the preexperimental procedure was to insure that (*a*) *S* had associated the poststimulus cues with their respective rows, (*b*) *S* was demonstrating "equal attention," i.e., all rows of the stimulus presentation were equally available on presentation of the poststimulus cue, and (*c*) practice effects would be minimal on the experimental day. Observation of the data on Day 3 and Day 4 verified that the preliminary training had produced the required effects.

The experimental protocol applied on the fourth day involved 108 presentations with every other presentation being the *same* slide. All other aspects of the procedure were exactly the same as on the previous days.

Each *S* was only scored if a digit was recalled in its correct position. Therefore, recalling all five digits of a given row in the correct sequence gave a recall score of 100%, where recalling only two in correct sequence gave a score of 40%.

Results and Discussion

Figure 1 shows the percentage of recall for each of the three delay intervals averaged across all *S*s. The recall-cue delay curve for the repeated slide does not differ from that for the nonrepeated series, an observation substantiated by the results of a Treatment (stimulus presentation, i.e., repeat or nonrepeat) × Treatment (delay of

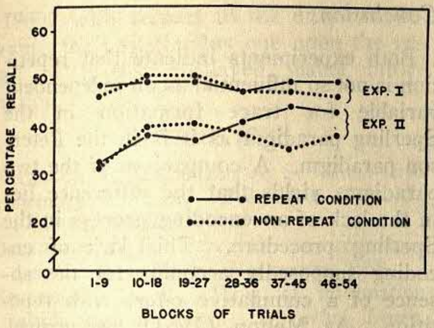


FIG. 2. Recall with respect to blocks of trials. (Each data point represents 72 observations.)

cue) × *S*s analysis. The *F* for the stimulus presentation condition was less than one. The MS_e for this effect was 19.94. The lack of an incremental effect with repetition is further demonstrated in Fig. 2, showing the percentage of recall across blocks of trials. With respect to the delay intervals the analysis yielded a significant effect with $F\ (2, 14) = 27.48$, $p < .001$, the MS_e being 54.61. All interactions were insignificant.

The results are in accordance with the assumption that the Sperling paradigm differs from the Peterson paradigm with respect to the effects of repetition. The proposed reason for this is the lack of encoding in the Sperling paradigm. The following represents the difference between the two procedures:

(Peterson paradigm)
Stimulus presentation → encoding—
interval → retrieval

(Sperling paradigm)
Stimulus presentation—interval
→ retrieval

If the data are interpreted within this framework they are supportive of both Melton (1963) and Broadbent (1963). Since repetition is regarded as being the important independent variable for storage then it may be posited that this study is not dealing with "memory traces" but "stimulus traces." Further in accord with Broadbent unless a stimulus is encoded it will not be subject to some form of permanent storage—thus a cumulative effect of repetition would not be evidenced. However the possibility of interference contributing to the present findings cannot be ruled out. Melton (1963) was able to eliminate completely the effects of repetition in a Peterson paradigm by the use of eight items between repetitions. Steffy and Eriksen (1965) have supposedly demonstrated that intrastimulus interference is markedly present in the Sperling-type procedure. Therefore the three types of interference—proactive, retroactive, and interitem—might be affecting these present data.

EXPERIMENT II

A second experiment was conducted to study the possible proactive inhibition and retroactive inhibition influence. Wickens, Born, and Allen (1963) have demonstrated that changing classes of material in a Peterson paradigm produces a release of inhibition. Therefore if inhibition had been generated in Exp. I by repeated presentation of a single class of items, then interpolating slides of a different class between presentations of the repeated slide should reduce, if not eliminate, such inhibition. Thus Exp. I was repeated except for certain modifications. Slides composed of three rows of five letters were interposed between the digit slides. The number of slide presentations on Days 2, 3, and 4 were 72, 108, and 108, respectively. Eight male *S*s were drawn from the same population as those *S*s used in Exp. I. The design was counterbalanced such that four *S*s received the repeat condition on Day 3 and the nonrepeat condition on Day 4, i.e., on Day 3 they received 54 individual letter slides and 54 repetitions of a randomly selected digit slide. The other 4 *S*s received the reverse order.

Results and Discussion

Figure 1 compares the repeat condition with the nonrepeat condition for the three delay intervals. As with Exp. I the effect of the presentation condition (i.e., repeat or nonrepeat) was insignificant, $F\ (1, 7) = 2.16$ with MS_e of 36.79. The delay effect was significant, $F\ (2, 14) = 8.37$, $p < .01$. The MS_e for this effect was 21.12.

As Fig. 2 indicates there is a gradual rise in the percentage of retrieval across blocks of trials. This is accounted for by the fact that half of the data constituting this graph were drawn from Day 3 and half from Day 4, thus allowing for practice effects to be evidenced. This factor may also account for some of the discrepancy between the data of Exp. I and the data of Exp. II observable in Fig. 1. A more noticeable difference between the two sets of data is the lack of deterioration in retrieval at the 700-msec. interval in Exp. II. If this phenomenon were due to the use of heterogeneous material then one might expect a similar retrieval-delay function to be obtained with the letter arrays. However this was not found to be the case. The retrieval-delay function for letters displayed a format similar to that obtained for the digits in Exp. I. Therefore, the apparent elevation of performance at the 700-msec. interval might be due merely to the flatness of the curve obtained, this in itself resulting from the overall lower level of performance generated by *S*s in Exp. II.

Congruent with Exp. I only two *S*s stated, when questioned, that they had been aware of repetition. They were, however, unable to recall more than three or four digits of the repeated slide and expressed that they had only become aware towards the end of the presentations.

Conclusions

Both experiments indicate that repetition is not so influential as an independent variable for trace formation in the Sperling paradigm as it is in the Peterson paradigm. A comparison of the two paradigms yields that the difference lies in the lack of an encoding process in the Sperling procedure. This lack of encoding supposedly accounts for the absence of a cumulative effect with repetition. As Melton (1963) has argued, the "preperceptual traces," as manifested in the Sperling (1960) and Averbach and Coriell (1961) studies, should be excluded from the domain of memory. For Melton the memory trace is only evidenced following a response to, or encoding of, the stimulus. Broadbent (1963) has made a similar observation by dichotomizing STM, one aspect dealing with stimuli which have not been "categorized." Therefore, in Melton's terms it might be proposed that mere repetition of a stimulus trace as opposed to a memory trace does not involve a structural change. It might be offered, however, that with this paradigm continued repetition would eventually produce a cumulative effect since the process of retrieval is somewhat similar to that of encoding, inasmuch that there is a translation of the visual display into a verbal form. This would, however, be dependent upon the correctness of retrieval.

A further theoretical implication of these experiments is that the storage system for stimulus traces is operationally different from the short-term storage system for memory traces. The model proposed by Broadbent (1958), which advocates a store for sensory data awaiting recognition (his S system) and obeying a temporal contingency, fits these data very well. However, although Exp. II rescinded somewhat the argument that proactive inhibition might be confounding the results, the possibility of very strong interitem interference still remains.

Finally, both experiments are offered as replicating the findings of Sperling

(1960) with respect to the effect of delaying the poststimulus cue upon the recall of briefly displayed visual information.

REFERENCES

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell Syst. Tech. J.*, 1961, **40**, 309–328.

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Broadbent, D. E. Flow of information within the organism. *J. verbal Learn. verbal Behav.*, 1963, **2**, 34–39.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* London and New York: Oxford Univer. Press., 1961. Pp. 37–46.

Hellyer, D. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *J. exp. Psychol.*, 1962, **64**, 650.

Melton, A. W. Implications of short-term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1–21.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

Steffy, R. A., & Eriksen, C. W. Short-term, perceptual-recognition memory for tachistoscopically presented nonsense forms. *J. exp. Psychol.*, 1965, **70**, 277–283.

Wickens, D. D., Born, D. G., & Allen, C. K. Proactive inhibition and item similarity in short-term memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 440–445.

(Received July 7, 1966)

(Continued from page 253)

Backward Conditioning: A Replication with Emphasis on Conceptualizations by the Subject: Arthur Zeiner and William W. Grings*: Department of Psychology, University of Southern California, University Park, Los Angeles, California 90007.

Comparison of Classical Conditioning and Relational Learning: Michael E. Dawson* and William W. Grings: Department of Psychology, University of Southern California, University Park, Los Angeles, California 90007.

Belief States and Sequential Evidence: Thornton B. Roby*: Applied Psychology Research Unit, Medical Research Council, 15 Chaucer Road, Cambridge, England.

Minimal and Maximal Sensory Intake and Exercise as Unconditioned Stimuli in Human Heart Rate Conditioning: Donald M. Wood* and Paul A. Obrist: Department of Psychiatry, The School of Medicine, University of North Carolina, Chapel Hill, North Carolina.

Trace Cue Position, Motivation, and Short-Term Memory: Delos D. Wickens* and C. Kenneth Simpson: Department of Psychology, Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

Trial and Error versus "Insightful" Problem Solving: Effects of Distraction, Additional Response Alternatives, and Longer Response Chains: Gary A. Davis,* Alice J. Train, and Mary E. Manske: University of Wisconsin, Research and Development Center for Learning and Re-Education, 1404 Regent Street, Madison, Wisconsin 53705.

* Asterisk indicates author for whom address is given.

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 294–299

DIFFERENTIAL DISCRIMINATION OF FREQUENCY OF CUTANEOUS MECHANICAL VIBRATION [1]

GENEVIEVE D. GOFF [2]

University of Virginia

Measurements of differential threshold (delta f) of a cutaneous mechanical vibratory stimulus are accurate only if concomitant changes in subjective intensity of the stimulus are eliminated. In this study, curves of equal subjective intensity were 1st obtained from 4 Ss at 2 intensity levels, 20 and 35 db. above the absolute threshold of the 100 cps standard. Delta f was then measured at 25, 50, 100, 150, and 200 cps, all amplitudes being within the equal intensity curve of the particular S. With intensity cues thus eliminated, delta f is small below 100 cps but then increases markedly. It is lower at the higher intensity level.

For almost half a century there have been attempts to use cutaneous vibration as a means of transmitting information. Gault and his colleagues tried to transmit the spoken word directly to the skin (e.g. Gault, 1924; Gault & Crane, 1928). According to Geldard (1957), the fundamental difficulty of such an approach lies in the inability to make sufficiently fine frequency discriminations in the range occupied by the strong components of speech sounds. A more accurate knowledge of frequency discrimination is fundamental for any further work in this direction.

Early attempts to measure differential thresholds for frequency failed to eliminate cues other than frequency, especially the cues arising from changes in subjective intensity occurring concomitantly with frequency changes. If *S* is to discriminate only differences in frequency, subjective intensity must be kept constant. Knudsen (1923) was the first to utilize electronic equipment permitting adequate control of intensity. He tried to control subjective intensity by increasing the stimulus intensities at all frequencies to 50 times the minimum threshold amplitude. However, curves of equal subjective intensity developed by Hugony (1935) showed that this did not insure equality of subjective intensity. Also, Knudsen's *S*s were not highly trained, and he did not use proper psychophysical techniques.

Later attempts (Joël, 1935; Roberts, 1932) were hampered by inadequate choice of frequencies and insufficient range of amplitude variation. Other reports citing differential frequency values (Békésy, 1962; Périlhou, 1947) do not give adequate descriptions of instrumentation and procedure.

Sherrick's (1954) apparatus incorporated two major improvements over previous studies. First, he eliminated the cues arising from the transients which occur with the onset and termination of the stimulus. Such transients give substantially smaller limens (Spector, 1954). Second, Sherrick attempted to control for the damping effects of the skin on the vibrator by calibrating his equipment with a piece of sponge rubber stimulating *S*'s finger. He equated subjective intensities as Hugony had done and studied the approximate limits of frequency discrimination. He considered his data insufficient but suggested that frequency discrimination is poor above 100 cps and relatively good below 100 cps.

---

[1] Adapted from a dissertation submitted to the Department of Psychology of University of Virginia in partial fulfillment of the requirements of the PhD degree. The writer is greatly indebted to John F. Hahn for his guidance and encouragement in the conduct of this research and to Frank A. Geldard for his suggestions in planning the research. This research was supported under Contract Nonr 474 (06) between the Office of Naval Research and the University of Virginia.

[2] Now at Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.

The present experiment was undertaken to study frequency discrimination at low frequency values. The apparatus, patterned after that of Sherrick, incorporated his improvement to eliminate transients and provided better vibrator performance and waveform over the frequency range utilized. Damping effects were controlled completely by determining amplitude empirically while the vibrator was stimulating *S*. Equal subjective intensity values were determined empirically at two intensity levels for each frequency to be studied.

METHOD AND RESULTS

Apparatus.—The output of a Hewlett-Packard low-frequency sine wave generator was led to a switching amplifier which permitted stimulus presentation without the switching transients which occur with mechanical devices. The switching amplifier output was sent to either of two attenuators through a power amplifier to a Goodman Model V47 vibrator. This vibrator is capable of excellent performance over the range of frequencies used. Stimuli were presented in pairs of 1-sec. signals separated by 2 sec., with a 4-sec. interval between pairs.

The *S* sat with his right arm on an armrest, his hand supported by a piece of plastic, and his index finger resting in a steady position across a 19-mm. diameter hole in the plastic. A contactor 6.5 mm. in diameter attached to the vibrator stimulated the finger through the hole. The vibrator was secured to one end of a balance arm and the pressure it exerted on the skin was held constant at slightly over 8 gm. by a counterbalancing weight.

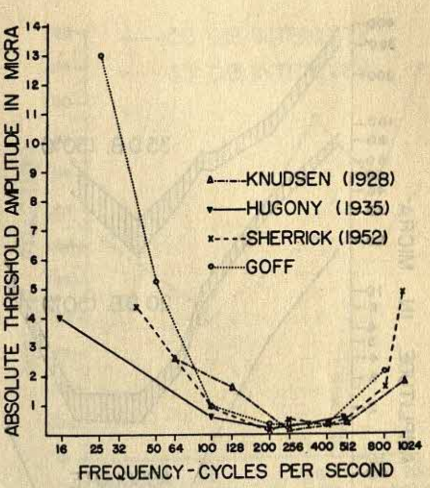


FIG. 1. Absolute thresholds as a function of frequency.

Earphones were used to provide a low frequency white masking noise, since audibility of the vibrator became a problem at the higher intensity level.

The amplitude of vibration was determined by a method first described by Geldard (1940). A needle attached to the contactor was observed through a microscope under stroboscopic illumination. If the frequency of the stroboscope is slightly different from the frequency at which the needle is vibrating, the movement of the needle is slowed and its amplitude may be accurately measured in micra. Since the amplitude may be somewhat damped by contact with *S*'s finger, and since the amount of this damping may vary from *S* to *S*, amplitude determinations were made under actual stimulating conditions for each of the four *S*s used, at both intensity levels of each frequency studied. These calibration curves did not vary from day to day.

Absolute thresholds.—After familiarizing *S* with vibratory stimuli, absolute thresholds were determined using the method of limits. On each of 3 different days, four ascending and four descending trials were run in balanced order at each of six frequencies—25, 50, 100, 200, 400, and 800 cps. Inter- and intrasession variability was minimal. Figure 1 shows that the mean thresholds of the four *S*s yield a U-shaped function similar to that reported by previous studies (cf. Hugony, 1935; Knudsen, 1928; Sherrick, 1952).

Measurement of equal subjective intensity

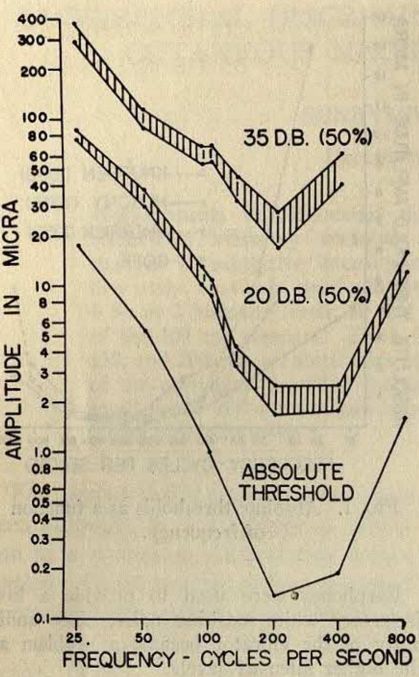


FIG. 2. Bands of equal subjective intensity for *S* BG.

—The *S*s were first given considerable practice in making comparative intensity judgments when presented with vibrations of different frequencies. The standard was 100 cps. Attenuation settings corresponding to 20 and 35 db. above threshold for the standard were computed for each *S* from the individual threshold measurements and calibration curves. Using the method of limits with equal judgments allowed, frequencies of 25, 50, 90, 200, 400, and 800 cps were compared randomly with the standard on 1 day, four ascending and four descending trials being run on each comparison. This procedure was followed on 3 days at the 20-db. level, and on 2 days at the 35-db. level. Similar sessions were run later with 130 and 150 cps when other points were found necessary to determine the shape of the curve. As a check on the equivalence of the two attenuator systems, the standard was also compared with itself at both intensity levels and the two systems gave substantially identical results. For three of the *S*s, 800 cps could not be compared with the standard at the 35-db. level since the necessary intensity could not be attained with the equipment.

Figure 2 shows the results of these equal intensity judgments for one *S*. The threshold curve is included in the figure for comparison. At each intensity level, the area between the lines encloses all those intensity values judged by *S* at least 50% of the time, on each day tested, as being equal to the intensity of the standard. Thus, this area contains all combinations of frequency and attenuation which are of equal "intensity" for *S*. The curves for the other three *S*s yielded similar "equal intensity bands."

As a check on the validity of the above method for finding curves of equal subjective intensity, the end points of the curves were then taken as standards. Thus, at each intensity level, the median intensity of the 25-cps stimulus which *S* had matched with the 100-cps standard was substituted as a new standard and 50 cps was matched to it. The range of intensities of 50 cps which had been chosen as equal to the 100-cps standard 50% of the time overlapped or was identical with the range of intensities chosen as equal to the new 25 cps standard 50% of the time. Both end points were matched in this manner for each *S* at each intensity level, with satisfactory results in every instance. Although *S*s claimed to be having more difficulty when comparing stimuli which were quite different in frequency, their intensity judgments were consistent.

Measurement of differential frequency.—The *S*s were next given practice in comparing vibrations on the basis of frequency. The standard was the frequency for which delta f was being found. Using the method of limits, an ascending trial began with the variable "equal" to the standard. Frequency was increased on each trial, attenuation being changed when necessary to keep in the center of *S*'s "equal intensity band," until the variable was reported as "faster" twice in succession. Descending trials began with the variable "faster" and continued until it was reported as "equal" twice in succession. Four ascending and four descending trials were run in balanced order for 25, 50, 100, 150, and 200 cps at both intensity levels on 2 different days. For values of more than 120 cps the variable frequency was changed in steps of 5 cps. For values of less than 120, the variable was changed in steps of .5 cps.

Delta f could not be found for 400 cps. When the variable was increased above this frequency, the intensity settings required were so high that distortion resulted. An attempt was made at 300 cps, but for only one of the four *S*s was delta f small enough that distorted waveform probably did not interfere.

The average delta f values of the four *S*s are plotted in Fig. 3, which shows that delta f is small at low frequencies but increases rapidly as frequency increases. Delta f is smaller at the higher intensity.

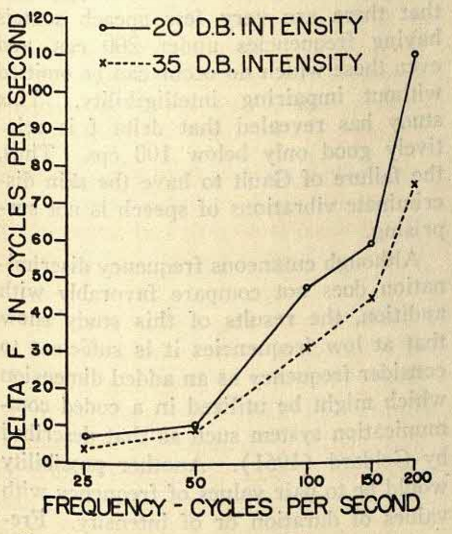


FIG. 3. Delta f as a function of frequency.

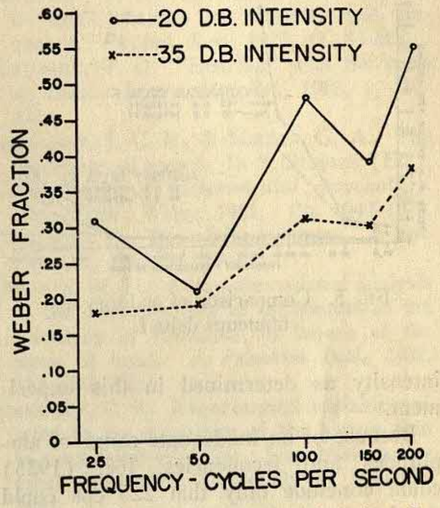


FIG. 4. Weber fractions as a function of frequency.

The mean Weber fractions of the four *S*s are plotted in Fig. 4. As frequency increases, the Weber fraction also increases, but more irregularly than does the corresponding delta f. The Weber fraction, like delta f, is lower at the higher intensity level.

DISCUSSION

Although earlier studies lacked adequate controls, their results can be compared with those of the present study. Since delta f decreases with an increase in intensity, the delta f values of Knudsen (1928) at 50 db. were expected to be lower than those of the present study at 35 db. His results are in the expected direction at low frequencies, but above 128 cps, his values are much lower than the other data would suggest. This probably means that his *S*s utilized inadequately controlled intensity cues.

Roberts (1932) reported that his *S*s could not distinguish 400 cps from 380, 410, or 420 cps. Figure 3 verifies that delta f at 400 cps is considerably more than 20 cps. In the cases where Roberts' *S*s could discriminate, an intensity cue was probably present since the variable was out of the zone of equal subjective

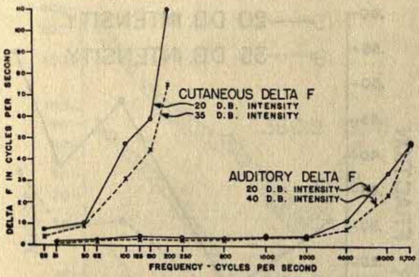


FIG. 5. Comparison of auditory and cutaneous delta f.

intensity as determined in this experiment.

Because of an inadequate range of amplitudes and frequencies, Joël (1935) could conclude only that 225 cps could not be discriminated from 200 cps and that 450 cps could not be discriminated from 400 cps. These results agree with the present study.

Périlhou (1947) studied the Weber fraction for frequency at 100 cps. He gave a value of 40% at "near threshold" intensity but did not report specific intensities. He states that delta f decreases regularly when intensity increases. Békésy (1962, Fig. 8) shows delta f increasing as intensity increases. This contradicts Périlhou's results, those of the present study, and results reported in audition. However, the discrepancy cannot be analyzed due to lack of detail of the Békésy and Périlhou methodology.

The equation $E = ka^2f^2$ (where k is a proportionality constant) expresses the relationship between energy, amplitude, and frequency in a vibratory system in which the amplitude is being varied sinusoidally. The possibility exists that energy differences could be a cue utilized in the reported frequency discriminations. As a check, the energies of the standard and comparison stimuli were computed. A detailed analysis of the results showed that large energy differences were sometimes present between stimuli which were not discriminable, and no energy differences were present between some stimuli which were reported as different. In the remaining cases, energy differences were not consistently in the same direction, but no reversals in the direction of the discrimination were reported. Therefore, it is improbable that energy could have been a discriminative cue for *S*s in this study.

Figure 5 compares the results of the present experiment with the data of Shower and Biddulph (1931) who studied delta f for audition over a wide range of intensities. The data of only two of their intensity levels are used in the comparisons, but they too found that delta f decreases as intensity increases. The difference in sensitivity between the two senses is striking. Only at frequencies below 50 cps does the skin even approach the sensitivity of the ear to frequency differences. Another difference is that in audition, delta f remains relatively constant up to the region of maximum absolute sensitivity of the ear (Sivian & White, 1933) and then rises. In fact, plotting the Weber ratios yields a U-shaped function with its minimum at the place of maximum sensitivity. In mechanical vibration, maximum absolute sensitivity is between 200 and 400 cps, but the delta f curves begin their sharp rise in the neighborhood of 50 cps.

Of specific interest for attempts to utilize the skin as a communication channel is the value of delta f for the range of frequencies contained in speech sounds. Licklider and Miller (1951) have noted that there are very few speech sounds having frequencies under 200 cps, and even those which do occur can be omitted without impairing intelligibility. This study has revealed that delta f is relatively good only below 100 cps. Thus, the failure of Gault to have the skin discriminate vibrations of speech is not surprising.

Although cutaneous frequency discrimination does not compare favorably with audition, the results of this study show that at low frequencies it is sufficient to consider frequency as an added dimension which might be utilized in a coded communication system such as that described by Geldard (1961). Another possibility would be to pair values of frequency with values of duration or of intensity. Frequency would then add redundancy to the system, as Howell (1958) suggested as a method of eliminating errors.

REFERENCES

Békésy, G. von. Can we feel the nervous discharges of the end organs during vibratory stimulation of the skin? *J. Acoust. Soc. Amer.*, 1962, **34**, 850–856.

Gault, R. H. Progress in experiments on tactual interpretation of oral speech. *J. abnorm. soc. Psychol.*, 1924, **19**, 155–159.

Gault, R. H., & Crane, G. W. Tactile patterns from certain vowel qualities instrumentally communicated from a speaker to a subject's fingers. *J. gen. Psychol.*, 1928, **1**, 353–359.

Geldard, F. A. The perception of mechanical vibration. *J. gen. Psychol.*, 1940, **22**, 243–308.

Geldard, F. A. Adventures in tactile literacy. *Amer. Psychologist*, 1957, **12**, 115–124.

Geldard, F. A. Cutaneous channels of communication. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 73–87.

Howell, W. C. Discrimination of rate of amplitude change in cutaneous vibration. Unpublished doctoral dissertation, University of Virginia, 1958.

Hugony, A. Ueber die Empfindung von Schwingunzen mittels des Tastsinnes. *Z. Biol.*, 1935, **96**, 548–553.

Joël, W. On the tactile perception of vibration frequencies. *Psychol. Rev.*, 1935, **42**, 267–273.

Knudsen, V. O. The sensibility of the ear to small changes of intensity and frequency. *Physiol. Rev.*, 1923, **21**, 84–102.

Knudsen, V. O. "Hearing" with the sense of touch. *J. gen. Psychol.*, 1928, **1**, 320–352.

Licklider, J. C. R., & Miller, G. A. The perception of speech. In S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1040–1074.

Périlhou, P. The vibratory sense. *J. gen. Psychol.*, 1947, **36**, 23–28.

Roberts, W. H. A two-dimensional analysis of the discrimination of differences in the frequency of vibrations by means of the sense of touch. *J. Franklin Inst.*, 1932, **213**, 283–311.

Sherrick, C. E. Experimental variables related to the sensitivity of the human skin to mechanical vibration. Unpublished doctoral dissertation, University of Virginia, 1952.

Sherrick, C. E. Measurement of delta f. Technical Report No. 24, 1954, University of Virginia, Project NR140-598, Office of Naval Research.

Shower, E. G., & Biddulph, R. Differential pitch sensitivity of the ear. *J. Acoust. Soc. Amer.*, 1931, **3**, 275–287.

Sivian, L. J., & White, S. D. On minimum audible sound fields. *J. Acoust. Soc. Amer.*, 1933, **4**, 288–321.

Spector, P. Cutaneous communications systems utilizing mechanical vibration. Unpublished doctoral dissertation, University of Virginia, 1954.

(Received July 11, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 300–302

SUPPLEMENTARY REPORTS

PACING OF REHEARSAL IN SEQUENTIAL SHORT-TERM MEMORY[1]

RICHARD A. MONTY, ROBERT KARSH

Supporting Research Laboratory, U. S. Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland

AND HARVEY A. TAUB

Psychology Service, Syracuse Veterans Administration Hospital

The Ss were required to keep track mentally of the number of occurrences of each of four different symbols presented sequentially. It was found that a light introduced into the sequence just prior to the presentation of each successive symbol tended to enhance performance. It was suggested that the light served to cue Ss to complete rehearsal of the current state of the information in preparation for the next stimulus in the sequence.

In a recent study (Monty, Taub, & Laughery, 1965), *S*s were required to tally mentally the number of occurrences of each of several different symbols (letters of the alphabet) presented visually in a sequence. In general, performance varied inversely with presentation rate, the number of different symbols presented in a given sequence, and the total number of symbols in a sequence. The results were explained primarily in terms of time available for rehearsal between successive symbols. In written reports, the majority of *S*s indicated that they found it most useful to picture a set of windows, one corresponding to each of the different letters presented. Individual running tallies were then kept in each window, and only the tallies were rehearsed after the presentation of each new symbol. A subsequent study by Monty, Karsh, and Taub (1967) further suggested that *S*s may attempt to perform multiple rehearsals of the tallies prior to the onset of the next signal and that without adequate cues for pacing, rehearsal may be disrupted by the onset of the next informative stimulus (i.e., stimuli which *S* must keep track of).

If interruption of the rehearsal process by the next symbol is a source of error in the keeping-track task, then it should be possible to employ a noninformative stimulus as a cue to alert *S*s to the pending conclusion of the off interval, enabling them to avoid disruption of the rehearsal process. Thus, one purpose of the present experiment was to determine if a light introduced in the sequence immediately prior to the presentation of each symbol would improve performance.

A second purpose was to determine if such factors as waning attention or increasing difficulty in estimating rehearsal time could account for the failure of successive increases in interstimulus off time to lead to concomitant improvements in performance with a short stimulus on time (Taub, Monty, & Laughery, 1967).

Method.—The *S*s were 90 enlisted men in the U. S. Army with a minimum General-Technical subtest score of 100 on the AGCT.

The apparatus has been described more fully elsewhere (Monty et al., 1965). Sequences of arabic numerals, seen as an outline of light against a dark background, were projected one at a time onto the surface of a 1½ × 1⅞ in. Industrial Electronic Engineers Incorporated rear-position display unit.

The *S*'s task was to observe and mentally tally the number of presentations of each of four different categories (the numerals 2, 3, 4, and 5) and to record the tally in writing at the end of each sequence.

Fifteen *S*s were randomly assigned to each of six groups derived from combinations of three interstimulus off times (2.2 sec., 4.2 sec., or 6.2 sec. between successive stimuli with on time held constant at 0.1 sec.) with two light conditions (No Light vs. Light). Thus, one group at each off time viewed a blank screen for the entire duration between successive symbols (No-Light condition) while the other was presented a green light of .3 sec. duration immediately preceding

---

[1] The authors gratefully acknowledge the assistance of Dennis F. Fisher and Richard L. Dalton in data reduction and analysis.

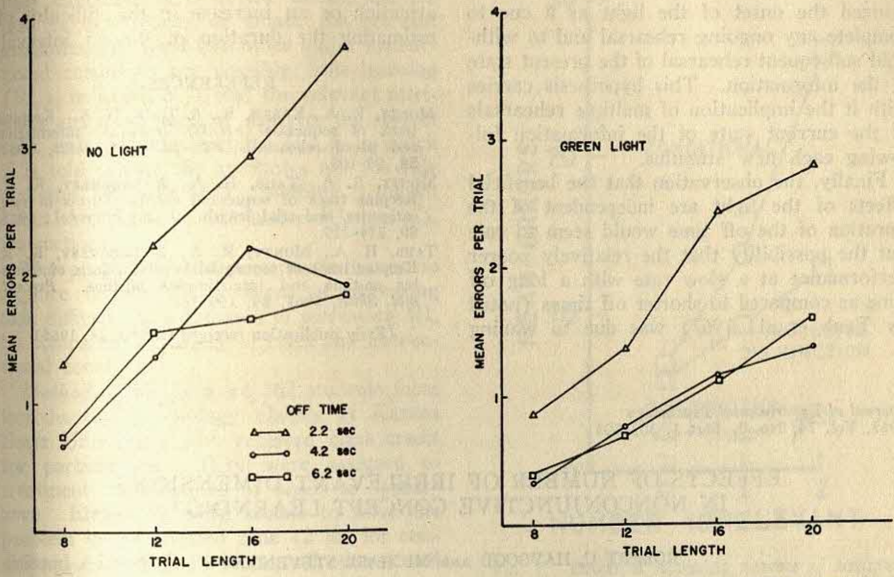


FIG. 1. Mean errors per trial as a function of off time and trial length for the no light and light conditions.

the onset of each symbol (Light condition). The light, which appeared on the display surface, occupied a portion of the interstimulus off time but did not lengthen the interval between successive symbols.

Trial length, defined as the total number of numerals presented in a sequence, was included as a within-*S*s variable. All *S*s were presented with Trial lengths of 8, 12, 16, and 20 symbols. To permit assessment of practice effects, each of the four Trial lengths was presented twice in random order during each of the four blocks of eight trials. The sequences of numbers were chosen at random with the restriction that each category appeared at least once on every trial.

The remaining procedures were identical with the previous study (Monty et al., 1965) except that a 2,500-hz. tone rather than a click was used to signal the beginning and end of each trial, and *S*s were given a 10-min. break between Trial 16 and 17.

Results and discussion.—Absolute error, defined as the sum of the absolute differences without regard to sign between the number of symbols presented in each category and the number reported, was used as the measure of performance (Monty et al., 1965). The absolute error scores for the two equal trial lengths within each block were summed and subjected to an analysis of variance. The data underlying the four significant main effects indicated that performance varied inversely with Trial length, F (3, 252) = 126.21, $p < .001$, directly with off time, F (2, 84) = 7.75, $p < .01$, directly with Blocks or Practice, F (3, 252) = 9.78, $p < .001$, and was generally better in the presence of the light than in its absence, F (1, 84) = 6.20, $p < .05$.

Only two of the interactions reached statistical significance. The data underlying the Trial Length × Blocks interaction, F (9, 756) = 12.66, $p < .001$, indicates as found in the earlier study (Monty et al., 1965) that although performance generally improved with practice, the amount of improvement varied directly with trial length. The trend of the Trial Length × Off Time interaction, F (6, 252) = 7.81, $p < .001$, illustrated in Fig. 1, and the observations that, in general, performance varied inversely with Trial Length and Rate of Presentation (i.e., directly with Off Time) are similarly in agreement with the previous results.

Of greater interest are the findings, as illustrated in Fig. 1, that (*a*) the introduction of a light for .3 sec. immediately prior to the onset of each successive symbol significantly improved performance as compared to the No-Light condition, and (*b*) the Light variable did not interact with the other variables but rather resulted in a general improvement in performance. It would appear then that *S*s were able to utilize the light to pace rehearsal. It is hypothesized that they utilized the onset of the light as a cue to complete any ongoing rehearsal and to withhold subsequent rehearsal of the present state of the information. This hypothesis carries with it the implication of multiple rehearsals of the current state of the information following each new stimulus.

Finally, the observation that the beneficial effects of the light are independent of the duration of the off time would seem to rule out the possibility that the relatively poorer performance at a slow rate with a long off time as compared to shorter off times (noted by Taub et al., 1967) was due to waning attention or an increase in the difficulty of estimating the duration of the off interval.

REFERENCES

MONTY, R. A., KARSH, R., & TAUB, H. A. Keeping track of sequential events: Irrelevant information and paced rehearsal. *Percept. mot. Skills*, 1967, **24**, 99–103.

MONTY, R. A., TAUB, H. A., & LAUGHERY, K. R. Keeping track of sequential events: Effects of rate, categories, and trial length. *J. exp. Psychol.*, 1965, **69**, 224–229.

TAUB, H. A., MONTY, R. A., & LAUGHERY, K. R. Keeping track of sequential events: Effects of stimulus on-time and interstimulus off-time. *Percept. mot. Skills*, 1967, **24**, 159–166.

(Early publication received October 24, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 2, Part 1, 302–304

EFFECTS OF NUMBER OF IRRELEVANT DIMENSIONS IN NONCONJUNCTIVE CONCEPT LEARNING[1]

ROBERT C. HAYGOOD AND MICHAEL STEVENSON

Kansas State University

The purpose of this study was to determine the relative effects of increasing the number of irrelevant stimulus dimensions under several different conceptual rules and instructional conditions. The results indicated that the effect of adding irrelevant stimulus information with nonconjunctive rules is similar to that previously found with conjunction, a linear decrement in performance. However, the rate of decrement is greater as rule difficulty increases. Similar results were obtained in a comparison of instructional conditions. The linear decrement with increases in irrelevant stimulus information thus appears to be quite general in concept learning, although the rate of decrement is highly dependent on specific characteristics of the conceptual problem.

Perhaps one of the best-established phenomena in concept learning is the increase in task difficulty as the number of irrelevant dimensions (NI) increases. Walker and Bourne (1961), using visually presented geometric patterns as stimuli, found a linear increase in errors and trials to solution as NI increased. Bulgarella and Archer (1962), using auditory signals, and Lordahl (1961), comparing auditory and visual stimulus information, also found essentially the same results.

These studies used either simple unidimensional (e.g., "red") or conjunctive (e.g., "red *and* square") concepts. Several investigators (Haygood & Bourne, 1965; Hunt, 1962; Neisser & Weene, 1962) have called attention to the existence of various types of nonconjunctive concepts, such as inclusive disjunction ("*either* red *or* square *or both*"), conditional ("*if* red *then* square"), and biconditional ("red *if and only if* square"). The first study designed to examine the effects of irrelevant information in nonconjunctive concept problems was that of Kepros and Bourne (1966). Using the biconditional rule, they found a linear increase in problem difficulty as NI increased. Because Kepros and Bourne used only biconditional problems, however, the relative influence of increasing NI with a variety of rules remains unknown.

All previous studies of NI have used instructions of the type described by Haygood and Bourne (1965) as attribute identification (AI). In AI, the conceptual rule is care-

---

[1] This research was supported in part by a research grant (MH 11283) from the National Institute of Mental Health, United States Public Health Service. It was conducted during the second author's tenure of a Traineeship sponsored by the National Aeronautics and Space Administration—Grant No. NsG (T)-54.

fully explained to S at the outset, and his task is to discover or attain the relevant attributes. At least two other major instructional conditions are possible: rule learning (RL), in which S is told the relevant attributes and must learn or discover the correct rule, and complete learning (CL), in which S is told neither the attributes nor the rule and must discover both. Nothing is known about the effects of NI under RL and CL instructions.

The purpose of the present study was to explore further the problem of increasing task difficulty as a function of increasing NI, by comparing a variety of rules and instructional conditions.

Method.—The *S*s were 162 students from introductory psychology classes at Kansas State University, who received class credit for participation. They were assigned to treatment combinations in order of appearance. Eleven *S*s were unable to solve the problem in the allotted time (2 hr. for conditional AI and CL, 1 hr. for all others) and were replaced. Seven of these failures to solve occurred in conditional AI and CL; the others were at the 1- and 2-irrelevant levels of the other rules.

The design of the experiment was a $3 \times 3 \times 3 \times 2$ factorial, with three levels of NI (0, 1, and 2 irrelevant dimensions), three conceptual rules (conjunction, inclusive disjunction, and conditional), three instructional conditions (RL, AI, and CL), and two pairs of relevant attributes (large-blue and one-square). The biconditional was omitted because of its extreme difficulty for the naive S under CL conditions.

The experimental task and apparatus were essentially the same as those described by Haygood and Bourne (1965). The *S*s were required to learn the correct method of sorting stimulus patterns (geometrical designs) to a criterion of 20 consecutive correct responses. After S was given detailed instructions appropriate to his instructional condition, the stimuli were rear projected on a screen one at a time. The S responded to each stimulus by pushing buttons labeled either YES or NO, which represented examples and nonexamples of the concept, respectively. Following each response, S was given feedback by the illumination of a signal lamp over the correct response button. The S was given as much time as needed to respond, and instructions stressed accuracy rather than speed.

Results and discussion.—In the analysis of variance of errors to criterion, three of the

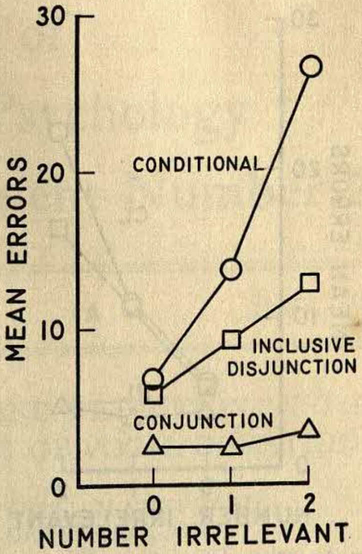


FIG. 1. Effects of increasing number of irrelevant dimensions for each of three conceptual rules. (Each plotted point represents the data from 18 *S*s.)

main effects were significant. Problem difficulty increased as NI increased, F (2, 108) $= 11.46$, $p < .01$. This function was essentially linear, $F_{\text{lin}} = 22.43$, $p < .01$, with the linear component accounting for 98% of the variance. The three conceptual rules differed in difficulty, with conditional the hardest and conjunction the easiest, F (2, 108) $= 21.70$, $p < .01$. Instructional condition also affected performance, with CL the hardest, AI intermediate, and RL the easiest, F (2, 108) $=$ 12.24, $p < .01$. Pair of relevant attributes was not significant. These results are the same as those found in earlier work (e.g., Haygood & Bourne, 1965; Neisser & Weene, 1962).

The crucial aspect of this study, however, was not these main effects. Rather, attention was directed toward two of the interactions. The Conceptual Rule × NI interaction, shown in Fig. 1, was significant, F (4, 108) $= 4.33$, $p < .01$. The effect of increasing NI was greater as rule difficulty increased, with conditional showing the largest effect and conjunction the least. Separate analyses of the three rules showed that the effect of NI was linear for conditional and inclusive disjunction (97 and 99% of the variance, respectively). The effect of NI was not significant for conjunction, which probably indicates that the task was too easy, and the design not sufficiently powerful, to

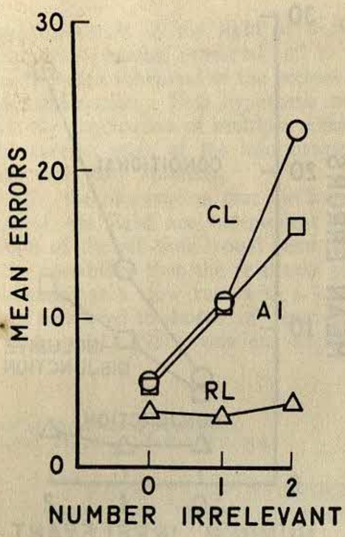


FIG. 2. Effects of increasing number of irrelevant dimensions for each of three instructional conditions. (Each plotted point represents the data from 18 *S*s.)

detect the relatively smaller influence of NI in conjunctive problems. The Instructions × NI interaction was marginally significant, $F(4, 108) = 2.98$, $p < .05$. This interaction is shown in Fig. 2. Pronounced linear effects of NI were found in separate analyses of CL and AI, accounting for 95 and 99% of the variance, respectively, but NI did not influence RL performance significantly. None of the other interactions was significant. Trials-to-criterion data showed the same general picture. However, because of relatively greater variability, the two important interactions failed to reach an acceptable level of significance.

These results indicate that the effect of increasing NI with nonconjunctive rules is similar to that previously found with conjunction, a linear decrement in performance. The rate of decrement increases, however, as the difficulty of the rule increases, which gives rise to an interaction of conceptual rule and NI. Similarly, the instructional condition affects the rate of decrement. Both CL and AI show strong effects, while the decrement in RL is minimal, as would be expected from the fact that in RL *S*s are told the relevant attributes at the outset. It appears that previous results showing linear increases in task difficulty from increases in irrelevant stimulus information are quite general, although the exact rate of decrement is highly dependent on specific characteristics of the conceptual problem.

REFERENCES

BULGARELLA, R., & ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *J. exp. Psychol.*, 1962, **63**, 254–257.

HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychol. Rev.*, 1965, **72**, 175–195.

HUNT, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

KEPROS, P. G., & BOURNE, L. E., JR. Identification of bioconditional concepts: Effects of number of relevant and irrelevant dimensions. *Canad. J. Psychol.*, 1966, **20**, 198–207.

LORDAHL, D. S. Concept identification using simultaneous auditory and visual signals. *J. exp. Psychol.*, 1961, **62**, 282–290.

NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *J. exp. Psychol.*, 1962, **64**, 640–645.

WALKER, C. M., & BOURNE, L. E. JR. The identification of concepts as a function of relevant and irrelevant information. *Amer. J. Psychol.*, 1961, **74**, 410–417.

(Received July 15, 1966)

Journal of
Experimental Psychology
Monograph Supplement Number 2

VOL. 74, PART 2 OF NO. 2 JUNE 1967

ACQUISITION, MAINTENANCE, AND RETENTION IN THE DIFFERENTIAL REINFORCEMENT OF VOCAL DURATION[1]

HARLAN LANE, JAMES KOPP, WILLIAM SHEPPARD,
THOMAS ANDERSON, AND DAVID CARLSON

Center for Research on Language and Language Behavior, University of Michigan

The durations of human vocal responses were systematically altered by selective reinforcement. The difference between the criterion duration for reinforcement and the mean duration of current responding, expressed in units of current response variability, $[(C\text{-}M)/\sigma]$ proved to be a major parameter of shaping. This "shaping index" correlated with the number of responses emitted before criterion was reached, and it was large whenever shaping failed. The duration of the ith response during shaping could be predicted as well from the mean of the 4 preceding response durations plus the mean of the 3 preceding increments in duration as from formulas involving more preceding responses or complex weighting functions. Newly shaped vocal durations retrogressed toward the base-line duration in the language when the speaker was returned to the language community for a day. However, there was marked retention, and reconditioning was accomplished much more quickly than the original shaping.

EXPERIMENT I

An excellent description of the procedure for differential reinforcement of responding has been given by Keller and Schoenfeld (1950):

> We select one (or more) of the "natural" variations of a well-conditioned response and give it *exclusive* reinforcement; the remaining variations are subjected to *extinction*. If we pick out, in advance, a variation that has been of fairly frequent occurrence, and if we apply this selective reinforcement rigorously, we can soon produce an increase in the frequency of the response that possesses the property . . . we have chosen. At the same time we decrease the frequency of those responses that do not meet our specifications.

In the experiments to be reported, the well-conditioned response that was selected was the emission of the vowel /u/ by human adults. The natural variants, given exclusive reinforcement, were durations of the vocal response greater than a criterion duration. The

---

[1] This research was supported in part by Contract OE-3-14-013 with the Language Development Section, United States Office of Education, and in part by Grant GB-552 from the National Science Foundation.

value of the criterion was an experimental variable; in several studies we picked out variants that were fairly frequent, in others, some that were not as frequent. Selective reinforcement was applied rigorously to every variant above criterion. There was an increase in the frequency of responses whose duration exceeded criterion and a decrease in the frequency of those variants that went unreinforced.

Method

Measurement of duration.—Techniques of measuring the parameters of the vocal operant have been discussed previously (Lane, 1960, 1964; Lane & Shinkman, 1963); the method employed to measure response duration will be summarized here. The vocal response was transduced by a dynamic microphone (Rye TM 1), amplified, and processed by an average speech-power circuit (Peterson & McKinney, 1961). The output waveform was applied to a dc amplifier (Krohn-Hite) and then to the time-interval section of a frequency counter (Hewlett-Packard 523D). The system was calibrated with pure tone pulses of known duration prepared with electronic timers and switches. The duration was read in milliseconds by the counter and recorded by a parallel printer and a card punch (IBM). This arrangement of apparatus was used in all the studies reported in this article.

Subjects and procedure.—Twenty male and eight female *S*s served individually in sessions lasting approximately 2 hr. The *S* was seated in an acoustically insulated chamber (Industrial Acoustics 400 A) in front of a panel containing a light and coin dispenser. These instructions were read to *S*:

> In this experiment you can earn money simply by saying /u/ when the light in front of you flashes. You will know that you have earned money when a penny falls into the tray; at the end of the experiment you may take home all the pennies you have earned. This is your only pay.
>
> Occasionally the light will not flash for a period of time. When the light is off, please do not respond.

An electronic timer flashed the stimulus light at 5-sec. intervals. After each of the first 20 responses emitted by *S*, *E* presented reinforcement and entered the response duration on a desk calculator. Then, *E* reset the timer and computed the reinforcement criterion for the next phase of the experiment. In the second and later phases, only those responses whose duration exceeded the criterion were reinforced. Each phase was continued until 10 reinforced responses occurred *in succession;* then the timer was reset, a new criterion computed, and the next phase initiated.

Criteria for differential reinforcement.—The duration criterion for reinforcement varied according to the schedule shown in Table 1. In Exp. IA, the criterion for each phase equalled the mean duration of the 10

TABLE 1

SUCCESSIVE CRITERIA FOR DIFFERENTIAL REINFORCEMENT OF VOCAL DURATION

Exp.	*S*s	Step Size	Phase Number								
			1	2	3	4	5	6	7	8	9
IA	1–10	+	CRF	M_1+ AD_1	M_2+[a] AD_2	M_3+ AD_3	M_4+ AD_4	M_5+ AD_5	M_6+ AD_6		
IB	11–15 & 20	+, +	CRF	M_1+ $.05M_1$	M_2+ $.10M_2$	M_3+ $.15M_3$	M_4+ $.20M_4$	M_5- $.05M_5$	M_6- $.10M_6$	M_7- $.15M_7$	M_8- $.20M_8$
IC	16–19	+, −	CRF	M_1+ $.05M_1$	M_2+ $.10M_2$	M_3+ $.15M_3$	M_4+ $.20M_4$	M_5- $.20M_5$	M_6- $.15M_6$	M_7- $.10M_7$	M_8- $.05M_8$
ID	21–23	Const.	CRF	M_1+ $.05M_1$	M_2+ $.05M_2$	M_3+ $.05M_3$	M_4+ $.05M_4$	M_5- $.05M_5$	M_6- $.05M_6$	M_7- $.05M_7$	
IE	24–25	Const.	CRF	M_1+ $.20M_1$	M_2+ $.20M_2$	M_3+ $.20M_3$					
IF	26–28	+	CRF	M_1- $.05M_1$	M_2- $.10M_2$	M_3- $.15M_3$					

[a] Mean duration of 10 terminal responses in Phase 2 plus their average deviation.

responses that terminated the preceding phase, plus their average deviation.

Following this experiment, the research detoured to obtain a preliminary estimate of the Weber fraction for produced vocal duration. Five additional college students were recruited for this purpose. These *S*s were instructed to produce 90 pairs of the vowel /u/ such that the responses in each pair had equal duration. Thirty pairs were obtained from each *S* at "short" durations, and a like number at "moderate" and "long" durations. When the absolute value of the difference in duration between the responses in each pair was plotted as a function of the shorter response, the 450 determinations were well fit by a straight line with a slope of 0.1 (determined by the method of least squares). It is interesting to note that the value of the Weber fraction, 10%, turned out to coincide with the average coefficient of variation (σ/M) for the duration of responses emitted under continuous reinforcement (see later).

In Exp. IB through IF the criterion for reinforcement in each phase equaled the mean duration of the 10 terminal responses in the preceding phase plus or minus some percentage of that mean. The step size or increment in the criterion ranged from half the value of the Weber fraction (5%) to twice that value (20%).

Probing for a tact of the reinforcement contingency.—At the end of each session in Exp. IB-IF, *S* was given a three-item questionnaire. A question was typed at the head of each of three pages. Handing the first page to *S*, *E* said: "I would like you to fill out this form completely." When *S* returned the first page, *E* gave him the second, and when he returned the second page he received the third. The questions were: (*a*) What do you think was the point of this experiment? (*b*) Did you know what it was about your responding that earned pennies for you? (*c*) Did you know that you received pennies depending on the length or duration of your response? If the words "duration" or "length" never appeared in *S*'s response and if the word "no" was part of the response to questions *b* and *c*, the questionnaire was scored "negative."

Results and Discussion

Experiment IA: Shaping longer durations; step size constant.—The relation between the criterion for differential reinforcement and the mean response duration in each phase is shown in Fig. 1 for a representative *S*. The terminal 10 reinforced responses, which ended each phase (filled triangles), the nonterminal reinforced responses (open triangles), and the unreinforced responses (inverted triangles) are plotted separately.

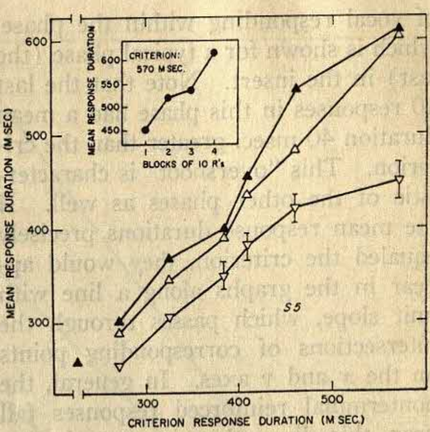


FIG. 1. Shaping longer vocal duration; step size constant, Exp. IA, *S*5. (The relation between the criterion response duration for differential reinforcement and the mean response duration. Unreinforced responses, inverted triangles, nonterminal reinforced responses, open triangles, and the 10 terminal reinforced responses, filled triangles, are plotted separately for each phase of the experiment. The base-line duration under continuous reinforcement is shown by the symbol at the lower left. The vertical lines represent 1 *SD*. The insert shows the mean duration of successive blocks of 10 responses in the last phase.)

It is apparent that the duration of vocal responding by *S*5 was under the control of the reinforcement contingencies. As the criterion was increased, in steps of one average deviation, from 270 to 570 msec., the mean duration of vocal responding was shaped from 258 to 610 msec. In each phase, the mean duration of unreinforced, nonterminal, and terminal reinforced responses increased in that order. This finding largely reflects the increasing duration of vocal responding within the phase, which is shown for a typical phase (the last) in the insert. Note that the last 10 responses in this phase had a mean duration 40 msec. greater than the criterion. This "overshoot" is characteristic of the other phases as well. If the mean response durations precisely equaled the criterion, they would appear in the graphs along a line with unit slope, which passes through the intersections of corresponding points on the x and y axes. In general, the nonterminal reinforced responses fall near this line, that is, at criterion values, while reinforced responses lie above, and unreinforced responses lie below the line.

Frequency distributions of response durations in each phase for *S*5 revealed that, under continuous reinforcement, response durations were short and the variance relatively small. With differential reinforcement, the variance increased and, in the later phases, the frequency distributions became bimodal. For most of the *S*s, the second peak lies at durations greater than the criterion and reflects the "overshoot" occurring toward the end of each phase, which was described above.

The frequency distributions also show than *S*5 reached the successive criteria rapidly; the number of unreinforced responses ranged from 5 in Phase 2 to a maximum of 20 in the last phase. A useful statistic to consider in this regard is the difference between the criterion duration and the mean duration of current responding, divided by the standard deviation of that mean: $(C\text{-}M)/\sigma$. This statistic, which we may conveniently call the shaping index, expresses the step size $(C\text{-}M)$ in terms of the variability (σ) of current behavior. An infrequent variant of behavior has been selected for differential reinforcement if the shaping index is large, either because the step size is large, or the variability is small, or both. Thus, the shaping index is an inverse measure of the operant level in the class of response topographies selected for reinforcement. For *S*s who served in Exp. I, the shaping index computed for each phase correlated .4 with the number of unreinforced responses in that phase (Spearman rho, $p < .01$). The shaping index for *S*5 was roughly constant at about .7; this follows from the fact that the criterion increment was one average deviation. The effect of the shaping index on the efficiency and effectiveness of shaping is examined in Exp. II.

Experiment IB: Shaping longer durations, step size increasing, then shorter durations, step size increasing.—Mean response duration as a function of criterion duration is shown for a representative *S* in Fig. 2. The mean duration grows steadily during the first five phases and the variability increases proportionally ($\sigma/M = .05$). In Phase 6, the criterion for reinforcement changed sign: any response duration *less than* 446 msec. was followed by reinforcement. The relative variability doubled ($\sigma/M = .10$); the mean duration of the unreinforced responses climbed abruptly to 570 msec.; while that of the terminal responses fell to 334 msec., an overshoot of 112 msec. The frequency distribution of response durations spread and became trimodal, with peaks corresponding, roughly, to the terminal, nonterminal reinforced, and unreinforced responses, respectively. Order was restored rapidly in the next phase, however, where the criterion increment was small and of the same sign. The shaping index for this *S* varied from 1.1 in Phase 2 to a maximum of 3.0 in the last phase. The latter value is large because the variability in the preceding phase was small

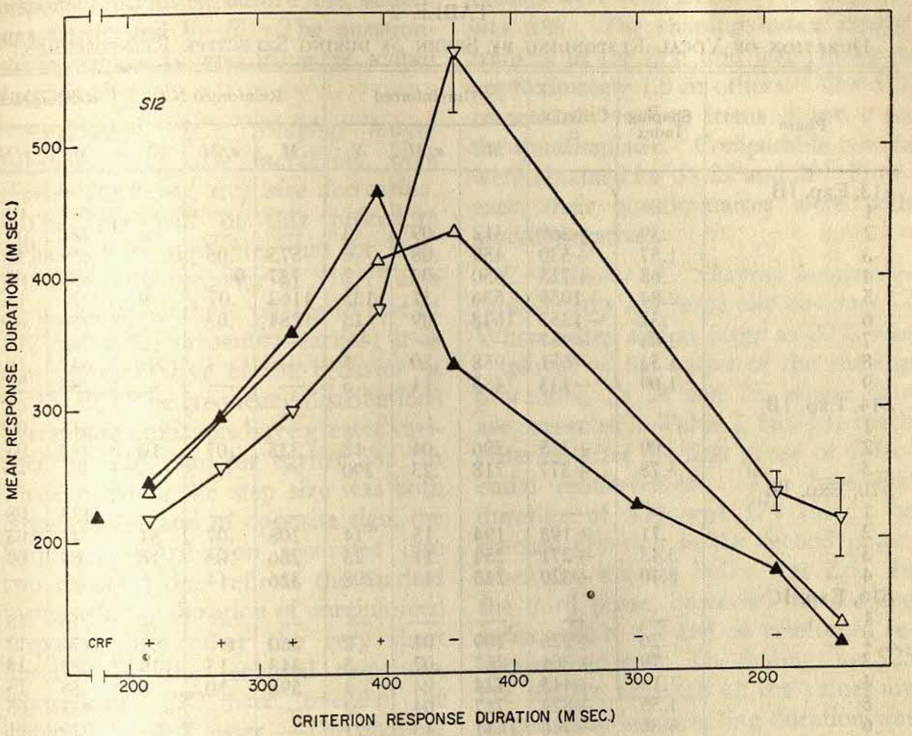


FIG. 2. Shaping longer durations, step size increasing, then shorter durations, step size increasing, Exp. IB, *S*12. (The relation between criterion and response duration. In Phase 6, the sign of the criterion was changed [−], and all responses with duration less than the criterion were reinforced.)

and the step size was 20% of the mean. There were nearly four times as many unreinforced responses in this phase as in any other. The questionnaire, which probed for a tact of the reinforcement contingency, was scored "negative" for this *S*.

The data for *S*13 (Table 2) provide another, comparable picture of the shaping of vocal duration. The criterion ranged upward from 369 to 1,353 msec. and then downward to 433 msec.; the mean duration of the terminal responses ranged from 348 to 1,424 to 285 msec. The largest shaping index (2.8) occurred in Phase 5 and there was a corresponding increase in the number of unreinforced responses (135, or 10 times more than in any other phase). When the sign of the criterion changed in Phase 6, an abrupt increase in the duration of unreinforced responses occurred. This finding is obtained whenever the sign of the criterion is changed: it may be related to the observation reported by Lane and Shinkman (1963), that several parameters of the vocal operant, including duration, increase in extinction after intermittent reinforcement. In all phases but the fifth, criterion was reached rapidly by *S*13, who described the reinforcement contingency on the first sheet of the questionnaire.

The findings obtained with *S*s who did not reach criterion in all phases are examined next (Table 2). The failure to shape the behavior of *S*14 may be

TABLE 2

DURATION OF VOCAL RESPONDING BY SEVEN *S*S DURING SELECTIVE REINFORCEMENT

Phase	Shaping Index	Criterion ±	Unreinforced			Reinforced NT			Reinforced T	
			M	σ/M	N	M	σ/M	N	M	σ/M
*S*13, Exp. IB										
1	—	—	—	—	—	—	—	—	348	.16
2	.39	+369	312	.09	3	—	—	—	483	.06
3	1.57	+530	459	.08	6	575	.05	3	659	.15
4	.68	+725	650	.09	2	727	0	1	883	.07
5	2.84	+1059	536	.31	135	1162	.07	9	1424	.14
6	.37	−1353	1648	.09	13	1284	.03	5	1186	.07
7	2.64	−967	—	—	—	—	—	—	766	.10
8	1.53	−651	958	.20	5	—	—	—	541	.18
9	1.09	−433	435	.14	3	—	—	—	285	.22
*S*14, Exp. IB										
1	—	—	—	—	—	—	—	—	313	.07
2	.09	+315	296	.04	12	335	.07	10	343	.03
3	3.78	+377	218	.23	290	—	—	—	—	—
*S*20, Exp. IB										
1	—	—	—	—	—	—	—	—	183	.08
2	.71	+193	194	.15	14	208	.07	31	247	.03
3	3.43	+271	254	.11	23	286	.03	10	287	.02
4	8.40	+329	243	.13	208	326	.14	5	—	—
*S*16, Exp. IC										
1	—	—	—	—	—	—	—	—	216	.18
2	.26	+226	209	.04	2	260	0	1	274	.12
3	.79	+301	267	.07	5	343	.13	10	388	.13
4	1.12	+445	424	.04	4	598	.10	8	688	.13
5	1.55	+824	717	.04	2	—	—	—	958	.05
6	4.36	−766	1631	.43	21	—	—	—	279	.21
7	.74	−236	306	.18	2	192	.04	4	156	.19
8	.45	−143	292	.59	5	116	0	1	122	.16
9	.05	−123	315	.74	3	—	—	—	59	.27
*S*21, Exp. ID										
1	—	—	—	—	—	—	—	—	238	.15
2	.36	+251	225	.08	6	272	.10	8	273	.05
3	.93	+287	266	.06	214	301	.04	83	332	.07
4	.70	+348	322	.06	47	367	.05	48	376	.04
5	1.19	+395	338	.14	5	—	—	—	438	.07
6	.76	−406	496	.04	5	361	.08	2	256	.11
7	1.00	−228	300	.20	11	172	.17	11	142	.12
*S*24, Exp. IE										
1	—	—	—	—	—	—	—	—	174	.07
2	2.62	+208	191	.05	28	218	.01	5	220	.04
3	4.89	+264	196	.11	197	—	—	—	—	—
*S*25, Exp. IE										
1	—	—	—	—	—	—	—	—	327	.13
2	1.23	+380	351	.12	45	399	.04	12	421	.06
3	3.81	+520	449	.11	239	528	.07	24	546	.04

Note.—Shown separately are measures of unreinforced responses, nonterminal reinforced responses, and terminal reinforced responses in each phase, under various values of the shaping index.

largely attributed to the size of the shaping index in the uncompleted phase: 3.8. It is interesting to note that the mean duration of the 290 unreinforced responses in the last phase fell *below* that mean for the preceding phase. The size of the shaping index similarly accounts for the breakdown of reinforcement control with *S*20. The step size in Phase 4 was moderately large (15%) and there was little variability in the preceding phase; hence, the shaping index took on the exceedingly large value, 8.4; 208 unreinforced responses occurred before the session was terminated by *E*. The questionnaires for *S*s 14 and 20 were scored "negative."

Experiment IC: Shaping longer durations, step size increasing, then shorter durations, step size decreasing.—The first half of this procedure (Phases 1–5) replicates that for Exp. IB. Thus the behavior of *S*16 (Table 2) is similar in these phases to that of *S*12 and *S*13, presented earlier; it is also representative of the behavior of *S*s 17–19. The frequency distributions of response durations have greater variance for *S*16 than for earlier *S*s. In Phase 6, when the step size was both large (20%) and of opposite sign, the frequency distribution separated into two clusters; one reflects the marked increase in the duration of unreinforced responses, the other the very short durations of the terminal responses, whose mean, 279 msec., overshot the criterion by 487 msec. The shaping index of 4.4 and the number of unreinforced responses were both larger for Phase 6 than for any other phase. The terminal response duration in Phases 7–9 was less than the mean duration observed under CRF; it reached as low as 59 msec. in the last phase, 157 msec. less than the base line determination. The *S*16 described the reinforcement contingency on the questionnaire.

Experiment ID: Shaping longer, then shorter, durations: step size small and constant.—When the shaping index remains small, as in the present procedure, variability remains small, shaping proceeds gradually and the number of steps required to change response parameters over a given range is increased. The *S* 21, for example, shows a steady increase in mean duration during the first five phases and then a rapid decline when shorter durations were reinforced. The step size was 5%. The shaping index ranged from .4 in the first and last phases to approximately 1.0 in others. The *S*21 responded "yes" to items *b* and *c* on the questionnaire. Comparable results were obtained for *S*s 22 and 23. However, their questionnaires were both scored negative.

Experiment IE: Shaping longer durations; step size large and constant.—When a step size as large as 20% was employed at the outset of the shaping procedure, *S*s 24 and 25, whose data are presented in Table 2, failed to reach criterion after the first phase of differential reinforcement. The base line duration of *S*24 was 174 msec.; he reached criterion in the second phase, when the shaping index was 2.6. In the third phase, however, the shaping index rose to 4.9 and no reinforced responses occurred. The findings for *S*25 are similar, although all the values are larger since his base line duration was 327 msec. The shaping index in the third phase was 3.8. A total of 273 responses were emitted in this phase, including 24 reinforced responses. It is interesting to note that the last 10 of these had a mean duration 26 msec. greater than criterion. However, numerous unreinforced responses were interspersed and the criterion of 10 successive responses with duration greater than 520 msec. was never reached. The questionnaire for *S*24 was scored negative whereas that for *S*25 was not.

Experiment IF: Shaping shorter durations; step size increasing.—The attempt to shape downward from the operant base line was successful only over a limited range. The findings for three *S*s are shown in Fig. 3. In Phases 2 and 3 the shaping indexes for all three *S*s ranged from .5 to 1.5 and differential reinforcement was effective.

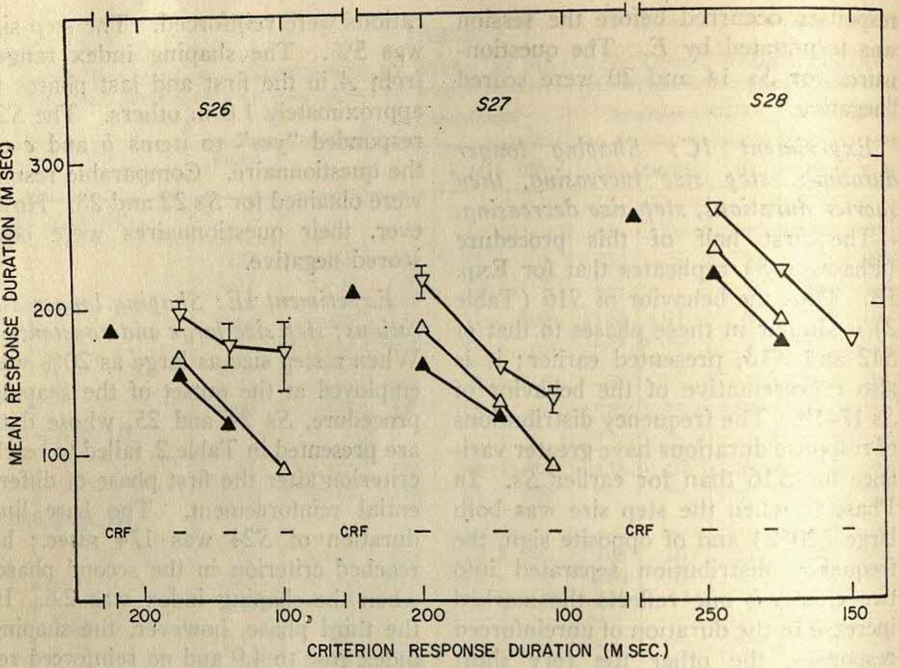


FIG. 3. Shaping shorter durations; step size increasing, Exp. IF, *S*s 26, 27, 28. (The relation between criterion and response duration. None of the *S*s reached criterion in the fourth phase.)

At the end of Phase 3, the mean duration and variability were quite small for all *S*s. As a result, the shaping index for the next phase was relatively large, approximately 2.5 for the three *S*s. No reinforced responses were emitted by *S*s 26 and 28 in Phase 4. Three responses by *S27* were reinforced but this did not lead to control over the duration of vocal responding. The questionnaires for all three *S*s were scored negative.

EXPERIMENT II

Experiment I showed that the topography of human vocal responses, like that of other operant behavior, may be altered by selective reinforcement. In order to manipulate vocal duration, we determined the initial frequency distribution of response durations (which was approximately normal), then selected a value of duration *n* standard deviations above the mean and finally reinforced all responses with durations longer than this criterion value. As a consequence, the mean and variance of the distribution of response durations increased. Complementary methods and findings were associated with the shaping of shorter durations.

Perhaps the most important parameter of the technique of shaping is the initial probability of those responses that have been selected for exclusive reinforcement. Keller and Schoenfeld (1950) advise, as stated earlier, that we "pick out, in advance, a variation that has been of fairly frequent occurrence"—but how frequent is "fairly?" As seen in Exp. I, if the initial probability of the responses selected for reinforcement is too low, there will be no shift in the distribution of response topographies. If the initial probability is quite high, on the other hand, the shaping process is secure, but large changes in topography can be accomplished only slowly and inefficiently. In quantitative terms, and taking vocal duration as a case in point, we want to know how many standard deviations (σ) away from the mean duration (M) of current responding should the reinforcement criterion (C) be set for effective and efficient shaping. In Exp. I we found that this parameter of shaping, $(C\text{-}M)/\sigma$, called the shaping index, was correlated with the number of responses that were emitted before criterion was reached and that it was large whenever the shaping process failed.

In this experiment, we systematically vary the size of the shaping index and observe its effect on the time course of shaping in detail. In addition, we are concerned with finding a simple way to predict the topography of responding at any moment in the shaping process, given the topography of the preceding responses.

Method

The experimental setting, the apparatus, the instructions to *S*s, the questionnaire, and the procedure for differential reinforcement duplicated those in Exp. I. Each of 39 college students served until either three consecutive shaping phases were completed or 2 hr. had elapsed.

The number of standard deviations above the mean, at which the criteria for reinforcement were set, varied across *S*s and across successive phases. The values employed were:

$$(C\text{-}M)/\sigma = .5,\ 1.0,\ 1.5,\ 2.0,\ 3.0.$$

The orders in which they were used are shown in Table 3. These values of the shaping index correspond, respectively, to the following probabilities of reinforcement (assuming a normal distribution of response durations):

.310, .160, .070, .020, and .005.

TABLE 3

VALUES OF THE SHAPING INDEX, $(C\text{-}M)/\sigma$, USED IN AN EXPERIMENT ON STEP SIZE AND THE CONTROL OF VOCAL DURATION

Exp.	Shaping Phase		
	1st	2nd	3rd
II A ($N = 6$)	.5	1.0	1.5
II B ($N = 10$)	1.0	1.5	2.0
II C ($N = 5$)	1.0	2.0	1.5
II D ($N = 5$)	1.5	1.0	2.0
II E ($N = 6$)	1.5	2.0	1.0
II F ($N = 7$)	2.0	3.0	4.0

Results and Discussion

Control of response topography.—Approximately half of the *S*s completed at least one shaping phase and did not tact the contingencies of reinforcement on the questionnaire. The mean duration and coefficient of variation (σ/M) in the phases completed by these 18 *S*s are given in Table 4.

TABLE 4

EFFECTS OF CRF AND SHAPING ON THE MEAN DURATION AND VARIABILITY OF VOCAL RESPONSES

	Phase			
	CRF $N = 18$	1st shaping $N = 18$	2nd shaping $N = 12$	3rd shaping $N = 8$
Mean Duration	236	264	309	430
Percentage of Increase from Duration in CRF		19	31	82
Coefficient of Variation	.10	.11	.11	.19

Note.—The data from phases completed first, second, and third were pooled across *S*s.

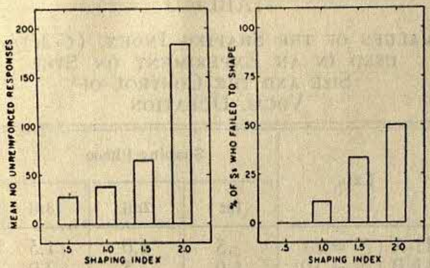


FIG. 4. Effect of step size on the efficiency of shaping. (The mean number of unreinforced responses during the first shaping phase and the percentage of *S*s who failed to complete that phase are shown as a function of the size of the shaping index: .5, 1.0, 1.5, and 2.0. The number of *S*s in each case is, respectively, 4, 9, 6, and 4.)

These results are comparable to those found in the earlier experiments cited above, with the mean duration increasing almost twofold in three shaping phases, and σ/M constant at about .10. The value of σ/M in the last shaping phase is higher than usual.

The effect of step size on the efficiency of shaping is shown in Fig. 4 in terms of the length of the first shaping phase when it was completed, and also the number of failures to complete it at all; the mean number of unreinforced responses and the percentage of *S*s who failed to shape are plotted as a function of the size of the shaping index. It is apparent that the lower the probability of occurrence of those responses selected for reinforcement, the longer the shaping process and the greater the danger of its unsuccessful outcome. The same relations were obtained, despite confounding with order effects, when the data from all *S*s were pooled over all phases.

The redistribution of response topography brought about by shaping is shown in detail in Fig. 5 for a representative *S*. Responses with duration less than criterion were sorted into two classes of approximately equal size and their frequency of occurrence in each class was plotted cumulatively as a function of consecutive responses. The same procedure was followed for reinforced responses. The cumulative record for responses with short durations, well below criterion, is negatively accelerated, whereas that for responses with durations well above criterion is positively accelerated. Typically, the intermediate functions, for class intervals bounded by the reinforcement criterion, have two points of inflection, indicating that the probability of the corresponding topography first increases at the expense of responses with shorter duration, and then decreases, in favor of responses with longer duration.

Prediction of response topography.— Several mathematical expressions were

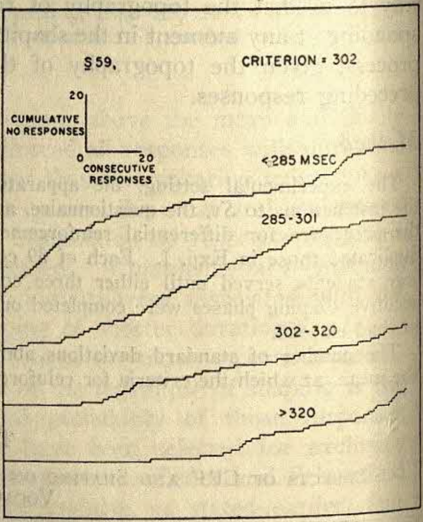


FIG. 5. Redistribution of response durations during shaping. (Response durations above and below criterion were sorted into two approximately equal classes and their frequency of occurrence in each class was plotted as a function of consecutive responses. The records have been displaced from their common origin to make them more legible. These data are representative of those obtained from 18 *S*s.)

compared as predictors of the topography of successive responses during shaping. In one quite general method employed for predicting the duration of Response $n+1$ from the preceding n response durations, all of the preceding durations were assigned weights and their mean computed (*a*); then the differences in duration between all pairs of consecutive responses were assigned weights and the weighted mean difference computed (*b*); finally, the sum of the weighted mean duration and the weighted mean difference was taken as the predicted duration (*c*);

$$(a)\quad \bar{X}_n = \sum_{j=1}^{n}(j+\gamma)^\alpha X_j \Big/ \sum_{j=1}^{n}(j+\gamma)^\alpha$$

$$(b)\quad \overline{\Delta X}_n = \sum_{j=1}^{n}(j+\gamma)^\alpha (X_j - X_{j-1}) \Big/ \sum_{j=1}^{n}(j+\gamma)^\alpha$$

$$(c)\quad \hat{X}_{n+1} = \bar{X}_n + \overline{\Delta X}_n$$

In this way, the topography of each response by *S* (except his first) was predicted from those preceding, and a product-moment correlation between the predicted and obtained response durations was computed for each pair of test values of the parameters γ and α.

The initial computations, performed on an IBM 7090 computer, yielded the highest predictive accuracy with weighting functions that had strong positive acceleration; in effect, very early responses in the shaping series were not entering the prediction of later responses at all. Consequently, a third parameter was added, a "look-back" variable, *L*, which specified the number of responses preceding Response $n+1$ entering in its prediction. When the computations were carried out with three parameters for all *S*s and phases, it was found that the predictive accuracy was not appreciably better when γ and α were permitted to vary than when they were set equal to zero. Under the latter condition, each predicted response duration is simply the mean of the *L* preceding response durations plus the mean of the *L*-1 preceding increments in duration.

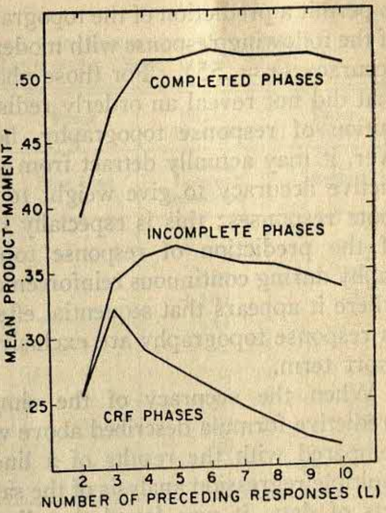


Fig. 6. Accuracy in predicting response topography during successful and unsuccessful shaping phases and CRF. (The average correlation between obtained response durations and those predicted from a mean-duration-plus-mean-difference formula is plotted as a function of the number of preceding responses entering the prediction of the next.)

The accuracy of predictions using the mean duration plus mean difference is plotted in Fig. 6 as a function of the number of preceding responses (*L*) entering the prediction. Correlations between predicted and obtained response durations, averaged over *S*s, are shown separately for shaping phases that were completed, for those that were not, and for CRF phases (where there was no progressive increase in response duration). It appears that, for those phases in which response topography was under experimental control, the prior four responses suffice to permit a prediction of the topography of the following response with moderate accuracy ($r = .55$). For those phases that did not reveal an orderly redistribution of response topography, however, it may actually detract from predictive accuracy to give weight to remote responses; this is especially true of the prediction of response topography during continuous reinforcement, where it appears that sequential effects in response topography are exclusively short term.

When the accuracy of the simple predictive formula described above was compared with the results of a linear multiple regression analysis of the same sets of data, it was found that linear regression accounted, on the average, for about the same amount of variance (16%–25%). The estimated partial correlations between each predicted duration and its predecessors X_n, X_{n-1}, . . . X_{n-9} considered singly were all close to $r = .10$, with the exception that the correlations between the duration of a response and that of its immediate predecessors were somewhat higher.

EXPERIMENT III

Experiment I demonstrated the control of response duration by differential reinforcement, while Exp. II inspected the time course of shaping, showed how it was affected by the major experimental parameter—the shaping index—and showed how it could be predicted using a simple mathematical expression. Experiment III takes cognizance of the fact that when selective reinforcement is applied—in the laboratory, the classroom or the clinic—to alter the vocal topography of an adult language user, the conditioning process competes during the period of acquisition with the speaker's long history of differential reinforcement for native-language responses, and it competes during maintenance with the speaker's continued successful use of native-language responses in his verbal community during the long intervals between conditioning sessions.

This experiment assesses the nature and degree of the deterioration of newly shaped vocal topography when the adult speaker was returned to his native-language community for a day. In particular, Exp. III sought to answer these questions:

(*a*) What is the test-retest reliability of the continuous reinforcement phase (CRF) as a measure of the average-response duration of individual *S*s prior to differential reinforcement (Exp. IIIA, IIIB)?

(*b*) Are durations measured during a CRF phase immediately following shaping comparable to durations measured at the end of the preceding shaping phase? In other words, does probing the differentiated repertoire under CRF alter the repertoire (Exp. IIIA)?

(*c*) Are responses of a given duration, shaped by differential reinforcement, retained after approximately 24 hr. in the language community (Exp. IIIB)?

(*d*) If response duration is shown to retrogress toward the language base line, is there nevertheless a reduction in the time required to reestablish the differentiated repertoire (Exp. IIIC, IIID)?

Method

The experimental setting, the apparatus, the instructions to *S*s, the procedure for differential reinforcement, and the questionnaire administered at the end of the experiment, duplicated those in Exp. I. The criterion for reinforcement in each shaping phase equaled the mean duration of the 10 terminal responses in the preceding phase plus the average deviation associated with that mean. Altogether 40 college students were employed, assigned at random to Exp. IIIA-D, 10 to each.

Table 5 shows that Groups IIIA and IIIB

TABLE 5

EFFECTS OF CRF, SHAPING, AND RETURN TO THE LANGUAGE COMMUNITY ON THE AVERAGE DURATION OF VOCAL RESPONSES (EXP. III A & B)

Experimental Treatment	Phase I	Phase II	Phase III	Phase IV–VII	Phase VIII
	Pretest 1	Pretest 2	Base Line	Shaping Longer Durations	Posttest
Group III A					
Day	1	2		2	2
Mean Duration	245	239		624	540
Percentage of Change		−2		+161	−13
Coeff. of Variation	.14	.10		.11	.11
Retention Index (%)					75
Group III B					
Day	1	2		2	3
Mean Duration	261	266		670	416
Percentage of Change		+2		+152	−38
Coeff. of Variation	.09	.09		.08	.11
Retention Index (%)					48

Note.—$N = 10$ for each group.

both received three phases of CRF, followed by four shaping phases and then a posttest under CRF. However, IIIB was returned to the language community for a day after the end of the last shaping phase and before the posttest.

Table 6 shows that Groups IIIC and IIID both received CRF, two shaping phases, return to the language community for a day, a posttest under CRF, and then shaping again. The latter reconditioning was continued for as many phases as it was possible to administer until *S* had emitted the same number of responses during retraining as he had during initial training. Experiments IIIC and IIID differed, however, in that successive criteria for reinforcement were gradually increased in the former by $C = M + \text{AD}$, whereas they were gradually decreased in the latter by $C = M - \text{AD}$. Thus, Group IIIC was shaped and later reconditioned to emit longer durations, whereas Group IIID was shaped and reconditioned to emit shorter vocal durations.

TABLE 6

EFFECTS OF SHAPING, RETURN TO THE LANGUAGE COMMUNITY, AND RECONDITIONING ON THE AVERAGE DURATION OF VOCAL RESPONSES (EXP. III C & D)

	Phase I	Phase II	Phase IV	Phase N
	Pretest (Day 1)	Shaping (Day 1)	Posttest (Day 2)	Reconditioning (Day 2)
Group III C (Shaping longer durations)				
Mean Duration	250	563	388	1634
Percentage of Change		+114	−31	+337
Coeff. of Variation	.12	.08	.13	.08
Retention Index (%)			44	
Savings (%)				28
Group III D (Shaping Shorter Durations)				
Mean Duration	239	167	203	156
Percentage of Change		+30	−22	+22
Coeff. of Variation	.11	.06	.09	.06
Retention Index (%)			39	
Savings (%)				−48

Results and Discussion

The changes in mean-response duration replication of the pretest and in shaping were quite comparable for Groups IIIA and IIIB (Table 5). The average durations in CRF-1 and CRF-2 (20 responses each) are virtually identical (±6 msec.) within groups, despite the fact that the test-retest interval was approximately 24 hr. Over the four shaping phases, the mean duration of the last 10 responses per phase increased to well over 600 msec. for both groups; relative variability remained nearly constant, indicating the differential reinforcement of longer duration led to an upward shift in the total response distribution. The average durations during the CRF phases agree well with the comparable measurements in Exp. I and II and with the value reported by Peterson and Lehiste (1960) for the average duration of /u/ as a syllable nucleus in American English. The coefficient of variation, ranging around 11% (±3%) agrees exactly with prior findings and with the initial estimate of the Weber fraction for produced vocal duration reported in Exp. I.

Retention.—When CRF conditions were instituted immediately after the fourth shaping phase, the average vocal duration decreased 13%. If the CRF posttest was delayed 24 hr., however, the retrogression toward the base line was 38%. To indicate what fraction of the ground gained by shaping had been lost, a "retention index" was computed as follows:

Retention in percent

$$= \left[1 - \frac{\text{Shape N} - \text{posttest}}{\text{Shape N} - \text{pretest}}\right] \times 100$$

where posttest = mean duration during CRF 3, pretest = mean duration during CRF 1, and Shape N = mean duration of the last 10 responses of the last complete shaping phase.

With no interpolated return to the language community, Group IIIA showed an average relative decrement of 25% or retention of 75%. The retention of the typical *S* was greater at the outset of the posttest and then quickly leveled off near the average value. This "drift" may be attributed to retrogression toward the first-language base line which occurs as soon as differential reinforcement is discontinued for a few responses, or it may be attributed to the operation of a "law of least effort"; the results for Groups IIIC and IIID, described below, clearly favor the former interpretation.

Group IIIB sustained an average relative loss of 52% (retention index = 48 %). Comparison with the findings for Group IIIA reveals that return to the language community causes a significant decrement in retention (Wilcoxon, 1949, rank test, $R = 74$, $p = .02$). Nevertheless, 48% retention can be considered substantial, since there were 24 hr. of intervening return to the language community (presumably involving differential reinforcement for short-duration responses), and as much as 25% loss is to be expected even when there is no interpolated delay (Group IIIA).

About 50% of the *S*s in Groups IIIA and IIIB correctly described the contingencies of reinforcement at the end of the experiment. Of the *S*s who did not, about half detected that topography of the vocal response was involved (pitch, vocality, etc.). A point-biserial correlation coefficient of +.66 ($p < .01$) was obtained between positive vs. negative questionnaire scores and mean response duration during the last shaping phase.

Turning to the findings for Groups IIIC and IIID, Table 6 shows that the base-line durations, the measures of variability, the change in duration effected by shaping, and the value of the retention index after interpolated delay are comparable (Wilcoxon rank test *ns*) for all groups in Exp. III (cf. Table 5). Moreover, *S*s whose vocal durations are shaped to longer values show a decrease of 31% toward the language base line, after the return to the community, or a retention of 44%, while *S*s whose vocal durations are shaped to shorter values show an *increase* of 22% toward the language base line, after the return to the community, or a comparable retention of 39%.

Once again, the statements based on averaged group data reflect individual performance well. Figure 7 plots the durations of responses emitted by *S*3 in Group IIIC. The systematic increase in duration during shaping, the retrogression following a 24-hr. delay, and the rapid reconditioning are all in evidence. However, *S*3—like others yielding comparable data—failed to describe on the questionnaire the contingencies of reinforcement; his responses to questions (*a*)–(*c*) were, respectively: "I don't know"; "No—maybe loudness"; "No." In order to explore the possibility that this *S* was correctly tacting an effective contingency of reinforcement during the experiment—albeit not the one planned by *E*—tape recordings of the experimental session were processed for intensity measurement (Lane & Shinkman, 1963). A plot of vocal intensity during the experimental session, analogous to that for vocal duration in Fig. 7, did not show a systematic increase. However, it should be acknowledged that the successful shaping of duration in other *S*s may have been influenced by the discrimination of an adventitious correlation between reinforcement and noncriterial parameters of vocal responding.

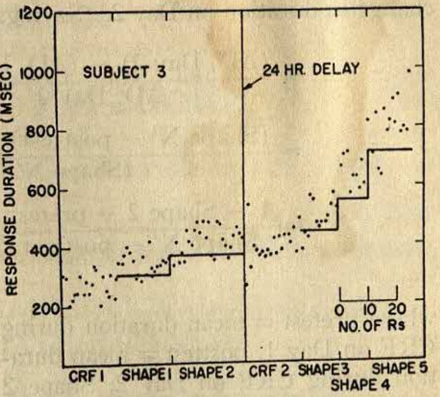


FIG. 7. Durations of successive responses emitted by an *S* in Group III C. (Response durations greater than criterion were reinforced. On the terminal questionnaire, this *S* failed to describe the contingencies of reinforcement.)

Savings.—In addition to the retention index, a second measure was obtained of the degree to which the effects of initial shaping were preserved despite return to the language community. This measure we call "Savings." Since the number of responses emitted, and hence the time spent, in reconditioning was set equal to that in initial shaping, the degree to which the change in duration during reconditioning exceeded that during initial shaping is a measure of the amount of savings. In particular, the extent of savings was taken to be the percentage of the total change in duration on Day 2 remaining after the absolute change in duration on Day 1 was subtracted from the absolute change in duration on Day 2. Savings in percent

$$= \frac{(\Delta\text{D, Day 2}) - (\Delta\text{D, Day 1})}{\Delta\text{D, Day 2}} \times 100$$

$$= \frac{(\text{Shape N}' - \text{posttest}) - (\text{Shape 2} - \text{pretest})}{(\text{Shape N}' - \text{posttest})} \times 100$$

$$= \frac{1 - \text{Shape 2} - \text{pretest}}{\text{Shape N}' - \text{posttest}} \times 100$$

where pretest = mean duration during CRF on Day 1, posttest = mean duration during CRF on Day 2, Shape 2 = mean duration of the last 10 responses of the second (last) shaping phase on Day 1, and Shape N′ = mean duration of the last 10 responses of the last complete shaping phase on Day 2.

Group IIIC had an average savings of 28% during shaping on Day 2 (Table 6). For Group IIID the savings index was −48%; this means that, during differential reinforcement of shorter response durations, there was less of a decrease from CRF to the last shaping phase on Day 2 than on Day 1. Response durations were probably approaching a lower bound in Exp. IIID. This is also indicated by the high inverse correlation (rho = −.63, $p < .05$) obtained between amount of retention and amount of savings for Group IIID contrasted with the zero correlation (rho = .02, *ns*) between retention and savings for Group IIIC. That is, in the case of shaping shorter durations, poor retention meant retrogression away from the area of minimal response duration and hence more room for reconditioning on Day 2, yielding, therefore, higher measures of savings.

Further evidence of a lower limit on vocal duration comes from point-biserial correlations for questionnaire scores and mean response duration during the terminal 10 responses of the last shaping phase; these were +.75 ($p < .01$) and +.09 (*ns*) for Groups IIIC and IIID, respectively. Apparently, discrimination of the contingencies of reinforcement enhanced shaping for the former Group but was nullified by a physiological limit in the latter Group. Seven *S*s in Group IIIC discriminated the contingencies of reinforcement, compared with only four *S*s in Group IIID. As in Exp. IIIA and IIIB, about half of the other *S*s detected that a vocal parameter was involved.

REFERENCES

KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.

LANE, H. L. Temporal and intensive properties of human vocal responding under a schedule of reinforcement. *J. exp. Anal. Behav.*, 1960, **3**, 183–192.

LANE, H. L. Differential reinforcement of vocal duration. *J. exp. Anal. Behav.*, 1964, **7**, 107–115.

LANE, H. L., & SHINKMAN, P. G. Methods and findings in an analysis of a vocal operant. *J. exp. Anal. Behav.*, 1963, **6**, 179–188.

PETERSON, G. E., & LEHISTE, I. Duration of syllable nuclei in English. Unpublished Report, Communication Sciences Laboratory, University of Michigan, 1960.

PETERSON, G. E., & MCKINNEY, N. P. The measurement of speech power. *Phonetica*, 1961, **7**, 65–84.

WILCOXON, F. *Some rapid approximate statistical procedures.* American Cyanamid Company, 1949.

(Received July 22, 1966)

Journal of
Experimental Psychology

VOL. 74, NO. 3 JULY 1967

DIRECTIONALITY AND ORGANIZATIONAL PROCESSES IN PAIRED-ASSOCIATE LEARNING[1]

MARLENE A. SEGAL AND GEORGE MANDLER

University of Toronto *University of California, San Diego*

A transfer paradigm was used with Ss 1st learning a paired-associate (PA) list and then a derived free-recall (FR) list, consisting of ½ the items from the PA list. For the 16-pair PA task, Ss learned either a standard unidirectional (A→B) list or a bidirectional (A↔B) list with pairs switching direction either systematically on alternate trials or unsystematically from trial to trial. Following PA learning, Ss were assigned to 1 of 4 FR conditions consisting of 16 items made up of (a) 8 pairs or (b) the vertical sets (A or B items) or (c) of a mixed set derived from the PA list. The FR data showed that in addition to the formation of pairwise horizontal associations, vertical organization also takes place during PA learning. Both vertical and horizontal-pairwise organization facilitated FR. The PA task indicated that unidirectional learning was generally better than bidirectional learning.

The free recall (FR) method has shown that *S*s' recall performance varies with the extent to which *S*s make use of interitem associations or coding devices (Bousfield, Puff, & Cowan, 1964; Tulving, 1962; Tulving & Pearlstone, 1966). Such organizational processes also take place during paired-associate (PA) learning. Battig (1966) has reported a series of experiments in which *S*s were instructed to recall both the A (stimulus) and B (response) items of each pair following each list presentation without restrictions on the recall order. When lists of heterogeneous difficulty were used, *S*s tended to group pairs in recall on the basis of their difficulty. Even with homogeneous lists *S*s tended to group pairs on the basis of their state of learning. Newly learned pairs were recalled at the beginning of recall while pairs learned earlier clustered later in recall. Battig also manipulated the presence of serial position, directionality, and difficulty cues in a PA list. PA learning was facilitated when these cues were present; grouping or coding of pairs in a PA list apparently reduced interpair interference.

The PA anticipation and recall methods may have concealed many such organizational processes, since *S* is re-

---

[1] The research reported here was supported in part by Grants GB 810 and GB 5282 from the National Science Foundation and APA 64 from the National Research Council, Canada.

quired merely to recall the response member of each pair when the stimulus item is presented. One process concealed in traditional studies is the formation of backward associations during forward PA learning, although their existence has been demonstrated with other designs (Asch & Ebenholtz, 1962; Feldman & Underwood, 1957; Jantz & Underwood, 1958; Murdock, 1956, 1958; Underwood & Keppel, 1963). Given that backward associations are formed, the following question still remains: When *S* learns an A-B association, does he simultaneously form a B-A association, *or* does he have to learn a new and separate association in the B→A direction?

Asch and Ebenholtz (1962) have presented the view that an association is bidirectional and symmetrical. They attempted to equate the availability of the A and B members of each pair, and found in most cases no difference in the probability of recall of forward and backward associations. Several other *E*s have handled the availability problem by using unidirectional and bidirectional lists, the latter requiring both forward and backward recall of each pair in the list (Schild & Battig, 1966; Underwood & Keppel, 1963), or double function lists (Battig & Koppenaal, 1965; Horowitz & Izawa, 1963), but many of their findings are inconsistent with Asch and Ebenholtz' results.

A transfer paradigm was used in this experiment with *S* first learning either a unidirectional or a bidirectional list and then an FR list consisting of selected sets of items from the *S*'s PA list. The purpose of the FR task was intended to uncover some of the organizational processes involved in PA learning. If organizational processes take place during PA learning that are not apparent in PA performance, they must be approached by transfer tests. The FR task is well suited for such a test since FR is particularly sensitive to organization defined as relations among members of a set of items.

METHOD

Subjects.—One hundred and eighteen University of Toronto freshmen participated in individual sessions. The data of 10 *S*s were eliminated because of apparent disinterest and subsequent failure to learn. The remaining 108 *S*s were randomly assigned to nine groups with the restriction that there were four male *S*s and eight female *S*s in each group.

Design.—The *S*s first learned a 16-pair PA list by the anticipation method and then a derived FR list consisting of half of the items from the PA list. For the first task, three conditions of PA learning were used. In Cond. PA-I, *S*s learned a standard *unidirectional* PA list with pairs being presented and recalled always in the same direction. The *S*s in Cond. PA-II learned a *systematic bidirectional* PA list with all pairs being presented and recalled in the A→B direction on Trial *n* and in the B→A direction on Trial $n + 1$. In this condition, the vertical sets of A items and B items were maintained over all trials. For Cond. PA-III, *S*s learned a *bidirectional* list, but half of the pairs switched direction *unsystematically* on each trial. For this condition there were no consistent vertical sets of A and B items.

Following PA learning, *S*s were assigned at random to one of four subgroups and were required to learn half of the items from the PA list using the FR method. The *S*s in Subgroup FR-P learned a list composed of half of the *pairs* (i.e., eight pairs from the PA list) presented one word at a time in random order. For *S*s in Subgroup FR-A, the list consisted of the *vertical set of A items* (the stimulus set) from the PA list. For Group FR-B, the FR list consisted of the *vertical set of B items* (the response set). The list for Subgroup FR-M consisted of a *mixed* list of one item from each A-B pair, such that half of the items were A items and half were B items.

Only three functionally different FR lists can be derived from the PA-II lists, since there were no distinctive stimulus or response sets for the systematic bidirectional lists. Similarly, because there were no consistent vertical sets in the unsystematic bidirectional lists, only two lists (a pairs list and a mixed list) could be derived from the PA-III lists. The design is an incomplete factorial design with nine functionally distinct cells and with 48 *S*s learning PA-I, 36 *S*s PA-II, and 24 *S*s PA-III.

Materials and list construction.—Thirty-two mono- or disyllabic nouns with AA frequency were selected from the Thorndike-Lorge (1944) tables, with the restriction that not more than two nouns started with the same letter of the alphabet.

Four FR lists were selected, and from these lists two PA lists were constructed for each of the nine conditions. Thus, all *S*s learned the same final FR lists. In constructing the PA lists, any primary associates such as "girl-friend" were eliminated. In each of the nine conditions, half the *S*s learned one PA list and half learned the other.

Four systematic list orders were constructed for the PA lists such that no pair appeared in the same serial position more than once, and no two pairs were adjacent more than once. Similarly, four different list orders were constructed for the FR lists. To prevent any systematic order effects, the presentation of list orders was counterbalanced among *S*s. A Stowe memory drum was used to present both PA and FR lists.

Procedure.—For PA learning, *S* was told that his task was to learn a list of 16 pairs of common words. The *E* explained that on each trial he would be presented first with one member of each pair and then he was to try to guess at the *missing* member of that pair before the whole pair would be shown to him. The same instructions were given to all *S*s regardless of which condition they were in. The *S*s were not informed in the PA instructions about subsequent FR learning.

Depending upon the group condition to which each *S* had been assigned, *S* was then presented with one of the PA lists. The pairs were presented at a 2:2-sec. rate for 16 trials, with a 4-sec. intertrial interval. The selection of 16 trials as the number of PA learning trials was based on the results from a pilot study where most *S*s reached the criterion of one perfect recall trial with a 16-pair list by the sixteenth trial.

Following PA learning, *S* was given a 2-min. rest and was then read the FR instructions. The *E* explained that *S* would now be required to learn half of the items selected from the PA list. He was told that following each presentation of the 16 items he was to recall aloud as many of the items as he could remember in any order that he wished. The appropriate FR list was then presented to *S* for eight trials, with items appearing one at a time at a rate of one word per sec. After each presentation *S* was given 45 sec. for recall. The *E* recorded all of *S*'s responses in the order recalled on each trial.

RESULTS

Paired-associate learning.—For the first analysis, the performance of *S*s in the three PA groups was compared by scoring the data in terms of the number of correct anticipations on each trial. Figure 1 shows that performance of *S*s in PA-I was slightly, but consistently, higher than that of *S*s in PA-II and PA-III, who learned the bidirectional lists. When the total performance of each *S* is summed over 16 trials, the mean number of correct responses for PA-I, PA-II, and PA-III, were 182.6, 167.0, and 159.4, respectively. A one-way analysis of variance showed a significant difference among the three groups, *F* (2, 105) = 5.38,

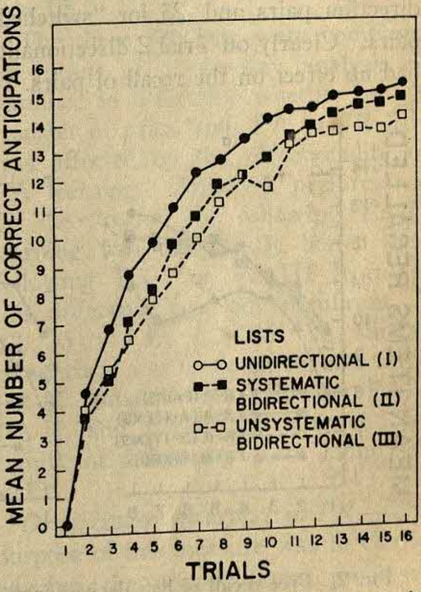


FIG. 1. Paired-associate learning curves for the three different PA lists in Task 1.

$p < .05$. *T* tests showed that performance on the unidirectional lists was significantly better than performance on both bidirectional lists—PA-I vs. PA-II: $t\ (83) = 2.33$, $p < .05$, and PA-I vs. PA-III: $t\ (71) = 3.19$, $p < .005$. Performance on the two bidirectional lists was not significantly different—PA-II vs. PA-III: $t < 1$.

To examine further bidirectional and unidirectional associations two more analyses were performed. On Trial 2, the first recall trial, the probability of correct recall of a pair was determined given that (*a*) it was to be recalled in the *same* direction as on Trial 1, or (*b*) it was to be recalled in the *opposite* direction. In Group PA-I, where all pairs remained in the same direction, the probability of correct recall was .29; in Group PA-II, where all pairs switched direction, the probability of correct recall was .24, $t\ (83) = 1.27$, $p > .05$. The corresponding probabilities for correct recall for the PA III condition were .24 for "same" direction pairs and .25 for "switched" pairs. Clearly on Trial 2 directionality had no effect on the recall of pairs.

The second analysis was prompted by remarks made by many *S*s in the bidirectional learning groups, that before they had learned a given pair, bidirectional learning was difficult, but once they had learned an association, the directionality of recall made no apparent difference. To test this suggestion, each individual pair for each *S* was traced from Trial 1 to its first correct recall on Trial *n*. Then Trial $n + 1$ was examined and it was noted whether or not the pair switched direction on Trial $n + 1$. It was found that the probability of being correct on Trial $n + 1$ was .75 for the unidirectional PA-I lists, and only .58 for the bidirectional PA-II lists, $t\ (83) = 5.10$, $p < .01$. Similarly, for PA-III, the probability of being correct on Trial $n + 1$ was .69 if the pair remained in the *same* direction and only .50 if the pair had to be recalled in the *opposite* direction, $p < .05$ by the comparison of counts test (Bennet & Franklin, 1954).

Free recall.—It was assumed that the greater the degree of organization for a set of items during PA learning, the better would be the performance

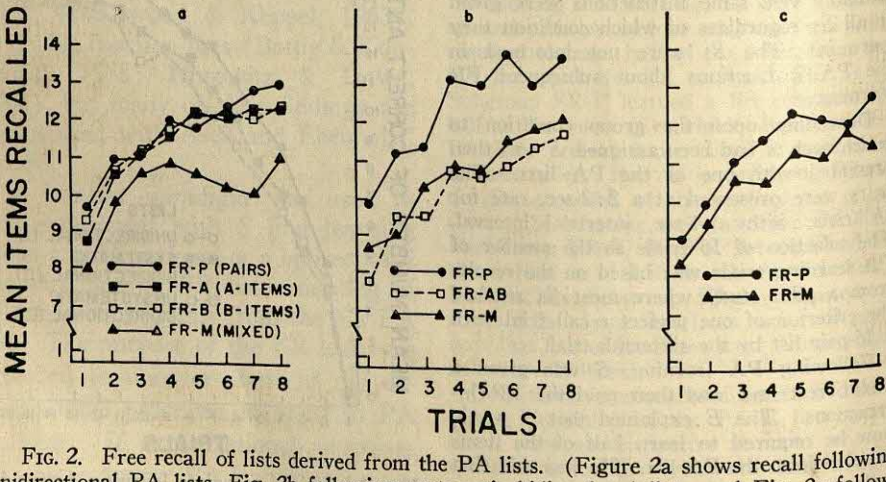


FIG. 2. Free recall of lists derived from the PA lists. (Figure 2a shows recall following unidirectional PA lists, Fig. 2b following systematic bidirectional lists, and Fig. 2c following unsystematic bidirectional lists.)

on the derived FR list consisting of that set. Figure 2 shows mean FR performance curves over the eight trials. The curves in Fig. 2a reflect the recall performance for the four FR lists of *S*s in PA-I; Fig. 2b, that of the three FR lists for *S*s in PA-II; and Fig. 2c, that of the two FR lists of PA-III *S*s; $n = 12$ for each group.

In all nine conditions there was relatively little improvement in performance from Trial 1 to Trial 8. The *S*s performance increased on an average by only three to four correctly recalled responses over all eight trials and reached asymptotic performance very quickly.

Figure 2a shows that for *S*s who had learned the unidirectional PA lists, performance on PA pairs (FR-P) and on both vertical sets (FR-A and FR-B) was about equal, whereas performance on the mixed item list (FR-M) was consistently poorer. An analysis of variance produced a significant difference among the FR lists, F ((3, 44) = 2.85, $p < .05$. The FR lists sum of squares was then subdivided by means of orthogonal comparisons in order to locate the significant differences among the four FR lists. This analysis clearly showed that nearly all the FR lists variance can be accounted for by the difference between the mixed item list and the three other FR lists, FR-M vs. FR-P, FR-A, FR-B: F (1, 44) = 8.01, $p < .01$. The "pairs" list and the two vertical item lists were not significantly different ($F < 1$).

Figure 2b shows the FR performance for *S*s in PA-II and demonstrates a consistently better performance of the FR-P list as compared to *both* the FR-AB and FR-M lists. A one-way analysis of variance indicated that performance on the three FR lists was significantly different, F (2, 33) = 9.15, $p < .01$, and a further breakdown of the FR lists sum of squares showed that performance was significantly better for the FR-P lists when compared with the two other lists, F (1, 33) = 16.25, $p < .01$.

Thus, some vertical organizational process took place during the learning of the unidirectional lists, but was apparently absent when the items switched sides on alternate trials. A t test between the means of the vertical lists for PA-I and PA-II, i.e., Lists FR-A + FR-B (PA-I) vs. FR-AB (PA-II), showed the difference to be significant, t (35) = 2.40, $p < .05$. Vertical transfer from a unidirectional list was better than from a bidirectional list.

Figure 2c shows the performance for the two FR lists for the PA-III bidirectional condition. Even though the recall of the FR-P list was consistently better than the recall of the mixed FR-M list this difference was not statistically significant, F (1, 22) = 1.91, $p > .05$.

The three FR lists were compared by means of a one-way analysis of variance to determine whether or not transfer of pairs from PA to FR lists was affected by the directionality of PA learning. Although performance on the "pairs" list following PA-II learning was consistently better than following PA-I or PA-III learning, this difference was not significant, F (2, 33) = 1.49, $p > .05$. A similar comparison was made among the three FR-M "mixed" lists following PA-I, PA-II, and PA-III learning and no significant differences were found ($F < 1$).

Organization measures in FR.—The purpose of this measure was to clarify whether the organizational processes in PA learning produced interitem associations. Tulving's (1962) subjective organization (SO) score was the measure used. SO is an estimate of the sequential redundancy in the recall protocols of a given *S* over *t* trials ($t = 16$ trials in the present experiment). The SO value was computed according to the formula:

$$\mathrm{SO} = \frac{\sum_{ij} n_{ij} \log n_{ij}}{\sum_{i} n_i \log n_i}$$

$$\text{such that } 0 \leqslant \mathrm{SO} \leqslant 1$$

where n_{ij} represents the frequency with which Word *i* was followed by Word *j* in *S*'s recall protocols, and where n_i represents the frequency with which Word *i* is recalled. $\Sigma n_i \log n_i$ is a measure of maximum organization, while $\Sigma_{ij} n_{ij} \log n_{ij}$ represents the actual organization. If *S*s produce clusters of items and if these associated items are recalled in close temporal contiguity on repeated trials a high SO score will result. Perfect organization (SO = 1.00) represents the case where the words that an *S* recalls are in the same order on each trial.

Table 1 shows the mean scores for the nine FR conditions. Across all three PA groups the SO scores for the FR-P pairs lists were higher than those for the vertical and mixed lists. Moreover, there is no systematic difference among the SO scores for the vertical and mixed lists. Statistical analyses of the SO scores showed that for Group PA-I, List FR-P was significantly higher than Lists FR-A, FR-B, and FR-M, $F\ (1,\ 44) = 15.00$, $p < .01$; for Group PA-II, List FR-P was significantly higher than Lists FR-AB and FR-M, $F\ (1,\ 33) = 91.19$, $p < .01$; but for PA-III, there was no significant difference between List FR-P and List FR-M, $F\ (1,\ 22) = 3.04$, $p > .05$.

TABLE 1

Mean SO Scores for Free-Recall Learning

	FR-P	FR-A	FR-B	FR-M
	pairs	A items	B items	mixed
P-A I	.35	.29	.29	.28
P-A II	.44	.26	.26	.28
P-A III	.34	.29	.29	.29

If the SO scores are calculated for alternate rather than for adjacent items in *S*'s FR output, all these differences disappear. Thus the organizational units for the pairs list (FR-P) is equal to two items.

Discussion

Evidence for coding processes.—In analyzing the FR data, it was assumed that the more readily a set of items could be retrieved from the PA list, the more organized the set of items must be. Compared with the mixed, unorganized FR lists, *S*s showed significant positive transfer both for the list consisting of pairs and for the lists consisting of vertical A and B items from the unidirectional PA lists, but only for the list of pairs from the systematic bidirectional PA lists. Thus *S*s organized the unidirectional lists both horizontally (in pairs) and vertically (in A and B sets), but the bidirectional lists were organized only horizontally. The lists of pairs (FR-P) also produced significantly higher SO scores, indicating that for the pairs lists pairwise associations were transferred from the PA to the FR lists.

What kind of vertical organizational process can occur in a unidirectional PA list but not in a bidirectional list? The low SO scores for all the vertical lists indicate that this vertical-organizational process probably did not consist of the formation of pairwise associations among the A or B items during PA learning. Rather the evidence suggests that some *coding* process took place during unidirectional learning that *S*s could not apply to the bidirectional lists. One of the main differences between the two PA conditions was that the vertical sets for the unidirectional lists always appeared on the same side from trial to trial, while the vertical sets in the bidirectional lists (PA-II) switched sides on alternate trials. The constancy of the position of the set affected its accessibility for recall.

The *S*s' reports substantiated this reasoning. All *S*s were asked at the conclusion of the experiment if they recognized how the 16 items comprising their FR list were selected from the PA list. For both vertical A and B item lists from PA-I, 7 out of 12 *S*s were able to state that their FR list consisted of either the right-side or left-side set from their PA list. In contrast to this, out of 12 *S*s learning the vertical sets from the bidirectional lists, only 2 *S*s reported that they thought the words always appeared together on the same side, χ^2 (1) = 4.45, $p = .05$.

These findings confirm the notion that *S*s who learned the unidirectional PA lists were provided with a "right-side" or a "left-side" cue which they could attach to each vertical set. No constant cues were available for the bidirectional (PA-II) vertical sets. The availability of these relevant cues seemed to aid *S*s in PA-I to retrieve the vertical sets during FR.

Directionality of associations.—On the basis of the PA results, one can neither fully accept Asch and Ebenholtz's (1962) principle of the symmetry of all associations, nor can one dismiss backward associations simply as examples of incidental learning as proposed by Feldman and Underwood (1957) and by Jantz and Underwood (1958). In this experiment the availability of the A and B items was maximized by the use of high frequency nouns, such that response learning of either the A or B items was minimized. On this basis, Asch and Ebenholtz should predict that there would be no difference in performance for unidirectional and bidirectional learning. The performance and the probability measures, however, showed that the learning of the unidirectional lists was in fact better than the learning of both the bidirectional lists. On the other hand, the finding that Trial 2 performance was independent of the direction of recall prevents one from rejecting outright the notion of the bidirectionality of associations.

Battig (1966) has suggested that the superiority in learning unidirectional over bidirectional lists may be due to the differential availability of directionality cues in the two kinds of lists. The results of this experiment support this view that coding processes and the use of cues play an important role in PA learning. The *S*s learning the unidirectional lists could code the 32 items into two sets of right-side and left-side items, and thus limit the response set to 16 items. This coding process cannot be applied readily to the bidirectional lists and, therefore, the response set for these lists is much larger. Thus it is possible that the poorer performance observed during bidirectional learning resulted from the lack of appropriate retrieval cues, rather than from the asymmetry or unidirectionality of associations. Associations may indeed be bidirectional but the unidirectional lists may provide *S* with more retrieval cues than do the bidirectional lists, thus concealing the bidirectional nature of associations.

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principle of associative symmetry. *Proc. Amer. phil. Soc.*, 1962, **106**, 135–163.

BATTIG, W. F. Evidence for coding processes in "rote" paired-associate learning. *J. verbal Learn. verbal Behav.*, 1966, **5**, 177–181.

BATTIG, W. F., & KOPPENAAL, R. J. Associative symmetry in S-R vs. R-S recall of double-function lists. *Psychol. Rep.*, 1965, **16**, 287–293.

BENNET, G. A., & FRANKLIN, N. L. *Statistical analysis.* New York: Wiley, 1954.

BOUSFIELD, W. A., PUFF, C. R., & COWAN, T. M. The development of constancies in sequential organization during repeated recall. *J. verbal Learn. verbal Behav.*, 1964, **3**, 489–495.

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

HOROWITZ, L. W., & IZAWA, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 352–361.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

MURDOCK, B. B., JR. "Backward" learning in paired-associates. *J. exp. Psychol.*, 1956, **51**, 213–215.

MURDOCK, B. B. JR. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

SCHILD, M. E., & BATTIG, W. F. Directionality in paired-associate learning. *J. verbal Learn. verbal Behav.*, 1966, **5**, 42–49.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teacher's College, Columbia University, 1944.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–354.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *J. verbal Learn. verbal Behav.*, 1966, **5**, 381–391.

UNDERWOOD, B. J., & KEPPEL, G. Bidirectional paired-associate learning. *Amer. J. Psychol.*, 1963, **76**, 470–474.

(Received July 23, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 313–320

CROSS-MODALITY MASKING FOR TOUCH AND HEARING[1]

GEORGE A. GESCHEIDER AND ROBERT K. NIBLETTE

Hamilton College

Auditory stimulation by a click was found to increase the tactile thresholds for mechanical pulses by as much as 5 db. The amount of auditory-tactile masking was related to the time interval between the click and pulse and the intensity of the click. Intense tactile pulses slightly increased the auditory click threshold, the maximum amount of masking being about 1.5 db. The amount of tactile-auditory masking was related to the time interval between the pulse and the click. When the ring and index fingertips of the same hand were stimulated by 2 pulses separated by a small time interval the probability of resolving 2 successive sensations was greatly reduced by presenting S with an intense auditory click. Furthermore, the reduction of temporal acuity of the fingertips was a function of the time interval between the auditory click and the tactile pulses.

Similarities between hearing and the skin senses have been the topic of much research in recent years (see Békésy, 1959). The observed similarities between touch and hearing suggest that the two modalities may interact in the perception of stimuli. When *S*s are required to make sensory judgments of alternating auditory and cutaneous stimuli, they often report that a pulse on the fingertip resembles a sound and, similarly, that a weak auditory click resembles a cutaneous tap. Furthermore, it has been observed that extraneous noises in the laboratory are often correlated with the fading or disappearance of tactile sensations. The above observations suggest that auditory stimulation may inhibit the perception of cutaneous stimuli and, conversely, that cutaneous stimulation may inhibit the perception of auditory stimuli. Consequently, the purpose of the experiments to be reported was to investigate (*a*) the effects of auditory stimulation upon tactile absolute thresholds, (*b*) the effects of tactile stimulation upon auditory absolute thresholds, and (*c*) the effects of auditory stimulation upon the temporal resolution of two tactile stimuli presented in rapid succession.

EXPERIMENT I

In studies of intramodality masking, the amount of masking is typically specified as a threshold increment: the difference between the thresholds measured in the presence of and in the absence of the masking stimulus. The intensity of the masking stimulus and the time interval between the masking and the test stimuli have been found to be major determinants of the size of the threshold increment. In the present experiment the threshold increment for the perception of brief taps on the fingertip was measured when the intensity of an auditory click and the time interval between the onset of the click and the onset of the tactile stimulus were varied.

Method

Subjects.—The *S*s were one member of the faculty and two students at Hamilton College.

Apparatus.—A set of Tektronix 160-series components consisting of a power supply, waveform generator, and two pulse gen-

---

[1] This research was supported by Public Health Service Research Grant No. MH 11966-01 from the National Institute of Mental Health.

erators, was used to produce 1-msec., square-wave, electrical pulses of variable time of onset. The output of the waveform generator, a negative-going saw-tooth wave, triggered the two pulse generators to emit two square-wave pulses at a rate of one pulse every .7 sec.

The pulses from one pulse generator were applied to a Goodmans V-47 vibrator after having passed through an impedance matching transformer and a Grason-Stadler graphic recording attenuator. The response of this vibrator to a 1-msec., square-wave pulse is an initial large positive displacement of the vibrator contact followed by a large negative displacement and a series of smaller positive and negative displacements (Sherrick, 1964). The duration of the major activity is approximately 10 msec., and a single pulse is felt as a brief tap. The *S* regulated stimulus-intensity level by pressing or releasing a hand switch. As long as he pressed the switch, intensity level decreased continuously at a rate of 1 db./sec. When he released the switch, intensity level increased at the same rate. The *S*'s manipulation of intensity level was recorded as a function of time by the graphic recording attenuator.

Pulses from the second pulse generator were applied to a Permoflux PDR-600 earphone. Intensity level was controlled by *E* through a decade attenuator box. The time interval between the onset of the auditory stimulus and the onset of the tactile stimulus was varied by manipulating the onset time controls of the two pulse generators.

Absolute thresholds for touch were determined for *S* in an Industrial Acoustics Company audiometric testing room. With his hand resting on his knee, *S* extended his index finger to make contact with the vibrator. To insure constant mechanical coupling between the vibrator and the fingertip, pressure of a ¼-in. diameter vibrator contact on the index fingertip was maintained relatively constant at 30 gm. by counterweighting a ball-bearing mounted shaft to which the vibrator was attached.

Procedure.—The *S*'s task was to track his touch threshold by continuously regulating the intensity level of the vibrator pulse. He was instructed to depress a switch until the taps on his fingertip disappeared, and then to release the switch until the taps became barely perceptible, at which time he was to depress the switch again until the taps disappeared. The presentation of a stimulus every .7 sec. was thought to eliminate, through habituation, the possibility that startle responses or other skeletal reflexes to intense auditory clicks could influence the touch threshold by altering the mechanical coupling between vibrator and fingertip. During the measurement of a single threshold *S* tracked for 1 min. and the threshold was estimated by drawing a line through the record of the up-down fluctuations in intensity as a function of time. To measure the masking effect of a particular auditory stimulus, *S* continuously tracked his touch threshold for a 3-min. period in which threshold shifts during auditory stimulation were observed. The touch thresholds measured during the first and third minutes served as base lines for evaluating the effect of presenting an auditory click with the tactile stimulus during the second minute. The amount that the touch threshold was masked by auditory stimulation was the difference between the touch threshold during auditory stimulation and the average of the touch thresholds measured during the two base line periods. The standard deviations for amount of masking obtained by this method were rather small, i.e., less than .5 db. in all cases where average masking was less than 1 db. and about 1 db. when average masking was 5 db.

During a session 10 such 3-min. threshold tracking periods were observed, one for each of the following time intervals between onsets of auditory and tactile stimuli: auditory stimulus preceding tactile stimulus by 45, 30, 20, 10, and 0 msec. (forward masking); and auditory stimulus following tactile stimulus by 45, 30, 20, 10, and 5 msec. (backward masking). The order in which thresholds for the 10 time-interval conditions were measured was random within each session. Eight such experimental sessions were conducted for each of four auditory-stimulus intensity levels (30-, 50-, 70-, and 90-db. SL). A different random order of the auditory stimulus intensities was employed for each of the three *S*s. In a random half of the eight sessions for each intensity condition the auditory stimulus was presented to the ear ipsilateral to the stimulated fingertip and in the other four sessions auditory stimulation was contralateral to the stimulated fingertip.

Results and Discussion

The amount of masking of tactile thresholds by auditory stimulation was at most 5 db. and occurred when a 90-db. sensation level (SL) auditory stimulus was presented simultaneously with the tactile stimulus. This effect, though slight, was very consistent both within and between *S*s. Moreover, as in intramodality masking the amount of masking was found to be systematically related to the time interval between the masker and the test stimulus and to the intensity level of the masker. In Fig. 1 the amount of masking is plotted as a function of the time interval between the onset of the auditory masker and the onset of the tactile stimulus (Δt). The four curves represent the relationship between Δt and the amount of auditory-tactile masking at four sensation levels of the auditory masker. Each point is the mean amount of masking for the combined data of the three *S*s who showed nearly the same amounts of masking at each Δt value. At all sensation levels maximum masking occurred when Δt was near zero. The amount of masking decreased sharply as the value of Δt increased for both forward masking (positive Δt values) and backward masking (negative Δt values). Comparable intramodality masking Δt functions for audition (Elliott, 1962; Raab, 1961) and for touch (Schmid, 1961; Sherrick, 1964) are steeper for backward masking than for forward masking. In the present experiment when the auditory masker was 90-db. SL masking decreased as Δt increased at about the same rate when the masker followed the tactile stimulus (backward masking) as when the masker preceded the tactile stimulus (forward masking). When the masker was 70-db. SL the Δt function was barely steeper for backward masking than for forward masking. This asymmetry of the Δt function becomes more pronounced at masker intensities of 50-db. SL and 30-db. SL.

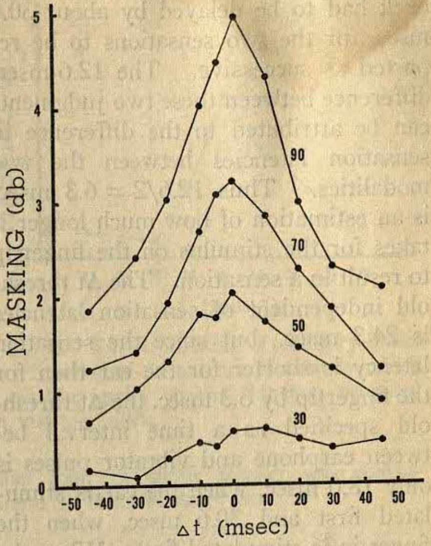


FIG. 1. Auditory masking of tactile pulses on the fingertip as a function of the amount of time by which the auditory masker preceded the tactile test stimulus. Sensation level of the auditory masker was either 30, 50, 70, or 90 db.

Neural impulses originating from weak taps on the fingertip probably reach the brain as long as 5–10 msec. later than neural impulses caused by simultaneously presenting an intense auditory stimulus. Therefore, the true Δt zero point in the present experiment is perhaps a −5 to −10 msec. time interval between the auditory and tactile stimuli. The hypothesis that sensation latency is shorter for intense stimulation of the ear than for weak stimulation of the fingertip was tested. When a 90-db. SL auditory click and a 10-db. SL pulse on the fingertip were presented simultaneously they were perceived as together in time. When the click was presented first and the pulse on the fingertip second they had to be separated in time by approximately 18.0 msec. to be perceived as successive. However, when the pulse on the fingertip was presented first the click had to be delayed by about 30.6 msec. for the two sensations to be reported as successive. The 12.6-msec. difference between these two judgments can be attributed to the difference in sensation latencies between the two modalities. Thus $12.6/2 = 6.3$ msec. is an estimation of how much longer it takes for the stimulus on the fingertip to result in a sensation. The Δt threshold independent of sensation latencies is 24.3 msec., but since the sensation latency is shorter for the ear than for the fingertip by 6.3 msec. the Δt threshold specified as a time interval between earphone and vibrator pulses is only 18.0 msec. when the ear is stimulated first and 30.6 msec. when the fingertip is stimulated first. When the Δt scale on the abscissa of Fig. 1 is corrected by adding a constant of 6.3 msec. to each Δt value, the results become consistent with those of intramodality masking experiments. The Δt function for a 90-db. SL masker is quite asymmetrical, with maximum masking occurring when the masker precedes the test stimulus by about 6 msec.

The difference between auditory and tactile sensation latencies should decrease as auditory transmission time is increased by lowering the intensity of the auditory stimulus. This could account for the tendency of the Δt functions in the present experiment to become more asymmetrical as the intensity of the auditory masker was reduced. Estimations of the difference between auditory and tactile sensation latencies obtained by the method described above for 70, 50, and 30-db. SL auditory clicks were 2.8, 1.5, and −1.3 msec., respectively. Each of the four masking curves replotted on a Δt scale, corrected for the difference between neural transmission times for auditory and tactile stimulation, shows approximately the same degree of asymmetry.

In contrast to the results of intramodality masking experiments, ipsilateral stimulation did not produce significantly more masking than contralateral stimulation. The amount of neural inhibition produced by masking within a sensory modality is highly correlated with the degree of neural interaction between stimulated areas. Two ipsilateral receptor areas may interact neurologically at both peripheral and central levels, but contralateral receptor areas interact neurologically only within the central nervous system. Thus intense stimulation of a fingertip will cause much larger threshold shifts for fingertips of the same hand than for fingertips of the opposite hand (Sherrick, 1964). It is highly unlikely that the specific auditory nervous system interacts either peripherally or centrally with the specific tactile nervous system and, therefore, the auditory-tactile masking observed in the present experiment could hardly have resulted from such neural interaction. Thus there is no reason to expect different amounts of ipsilateral and contralateral auditory-tactile masking.

A neural mechanism that might account for the suppression of sensations in one modality by stimulation in another is the reticular formation of the brain stem; a region where neural impulses from all sensory modalities converge. The activation of the reticular formation by sensory stimulation can result in inhibition of sensory impulses at every synapse along a sensory pathway. Inhibition has been demonstrated both within and between sense modalities (Hernández-Peón, 1961). Intense auditory stimulation activating the reticular formation is likely to result in inhibition of the perception of a simultaneous weak tactile stimulus.

EXPERIMENT II

In Exp. II tactile masking of the auditory perception of clicks was investigated. Preliminary observations revealed that tactile stimulation had very little, if any, masking effect on the absolute threshold for auditory clicks (less than 1 db.). Therefore, to determine whether tactile-auditory masking occurs at all, a relatively intense tactile stimulus (40-db. SL) was used throughout the experiment. As in Exp. I the time interval between the masking stimulus and test stimulus was systematically varied.

Method

Subjects.—The three *S*s of Exp. I served in Exp. II.

Apparatus.—The apparatus was the same as that used in Exp. I except that the graphic recording attenuator was used with the auditory rather than with the tactile stimuli. The intensity of the vibrator pulses was 40-db. SL, and *S* regulated the intensity of the auditory clicks as he tracked his threshold by manipulating the hand switch of the graphic recording attenuator.

It was noticed that even while wearing earphones the relatively intense vibrator pulses could be faintly heard. To eliminate the possibility of the sound of the vibrator pulses masking earphone clicks, 70 db. of white noise, measured a few inches in front of *S*'s head, was continuously present during each session. Under these conditions shifts in auditory thresholds during tactile stimulation could clearly be attributed to intermodality rather than to intramodality masking caused by sounds from the vibrators.

Procedure.—The procedure for measuring threshold shifts was the same as that used in Exp. I. During a session, threshold shifts were measured in a random order when the tactile pulse preceded the auditory click by 40, 20, 10, and 0 msec. (forward masking) and when the tactile pulse followed the auditory pulse by 40, 20, and 10 msec. (backward masking). From each *S* eight threshold shifts were obtained for each of the seven time intervals. In four threshold measurements the tactile masker was ipsilateral to the stimulated ear and in the other four the tactile masker was contralateral to the stimulated ear.

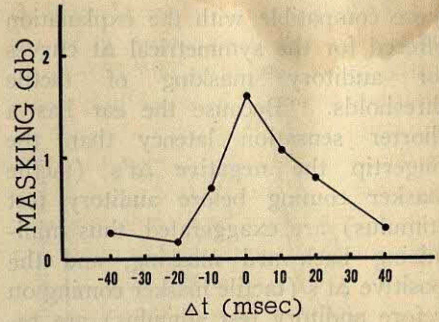


FIG. 2. Tactile masking of auditory clicks as a function of the amount of time by which a masking pulse on the fingertip preceded an auditory click.

Results and Discussion

Tactile stimulation had a very slight masking effect upon auditory click thresholds. The amount of tactile-auditory masking was at most 1.6 db., compared with 5.0 db. for auditory-tactile masking. A 40-db. SL tactile stimulus is more than moderately intense and is capable of arousing very strong sensations of touch. Thus it is not likely that the difference between auditory and tactile masking is due to a difference in masker intensity. This suggests that the auditory system has somewhat of a neurological dominance over the touch system.

In contrast with the results of the auditory-tactile experiment, the amount of tactile-auditory masking was considerably different for the three *S*s. Although the amount of masking was .6, 1.0, and 3.0 db. at Δt of zero for the three *S*s, the three Δt curves were combined since they were very similar in form. The resulting Δt curve (Fig. 2) is very similar in shape to those characteristic of intramodality masking, the time course of backward masking being more rapid than the time course of forward masking. This result is quite compatible with the explanation offered for the symmetrical Δt curves for auditory masking of tactile thresholds. Because the ear has a shorter sensation latency than the fingertip the negative Δt's (tactile masker coming before auditory test stimulus) are exaggerated, thus minimizing backward masking, and the positive Δt's (tactile masker coming on before auditory test stimulus) are reduced, thereby maximizing forward masking.

As was the case with auditory-tactile masking the amount of masking for contralateral and ipsilateral stimulation was the same, further supporting the idea that a central filtering mechanism such as the reticular formation mediates intermodality masking.

EXPERIMENT III

The time interval between the onsets of two brief stimuli necessary for the perception of two successive sensations has been used as a measure of auditory and tactile temporal acuity (Gescheider, 1966). Experiment III was designed to investigate the effects of intense auditory stimulation upon the temporal acuity of the skin. The ring and index fingertips of the right hand were successively stimulated by two brief mechanical pulses separated by a small time interval. The probability of *S*'s reporting successive sensations was determined as a function of the time interval between the onsets of an auditory click and the first of two successive tactile stimuli.

Method

Subjects.—One member of the faculty and two students at Hamilton College served as *S*s.

Apparatus.—In the audiometric testing room two vibrators were mounted on the same counterweighted shaft with the contacts placed 3 in. apart. In this arrangement *S* could easily place his right index fingertip on one vibrator and his right ring finger on the other. One-millisecond, electrical, square-wave pulses from two pulse generators, after passing through impedance matching transformers, were applied to the vibrators. Attenuators preceding each vibrator were set so that the sensation level of the two vibratory stimuli was 20 db., a value producing a moderately intense tactile sensation. The time interval between the mechanical responses of the two vibrators could be varied. Intense 90-db. SL auditory clicks were produced by applying 1-msec., electrical pulses from a third pulse generator to a Permoflux PDR-600 earphone. The time interval between the auditory clicks and the vibrator pulse could be systematically varied.

Procedure.—Two pulses were presented to *S* through the vibrators, and his task was to report whether he felt one or two taps. He was instructed to report "two" when sensations on the ring and index fingertips of his right hand were temporally separated and to report "one" when only one stimulus could be perceived or when both stimuli were perceived as simultaneous sensations. The time interval between the onsets of the vibrator pulses was either 31.0 or 38.5 msec., while the time interval between the auditory stimulus and the first of the two vibratory stimuli was varied over a range of 235 msec. The auditory stimulus preceded the first vibratory stimulus by 70, 51, 33, 26, 16, 9, 4, and 0 msec. and followed the first vibratory stimulus by 4, 8, 12, 20, 24, 30, 36, 41, 51, 62, 82, 103, 127, 147, and 165 msec. During each 1-hr. session stimuli were presented once every 5 sec. and several judgments were obtained for each of the auditory-tactile, time-interval conditions presented in a random order. Judgments were obtained within a session when the vibrator pulses were both 31.0 and 38.5 msec. apart. During the experiment for both the 31.0- and 38.5-msec. tactile intervals a total of 216 judgments was made for each of the 22 auditory-tactile, time-interval conditions and 432 judgments were made in the absence of the auditory stimulus.

Results and Discussion

The temporal resolution of tactile stimuli separated by short time intervals was found to be greatly affected by the presentation of an intense auditory click. The proportion of times that *S*s reported a single tactile sensation in the absence of the auditory stimulus was only .083, and .065 when the time intervals between vibrator pulses were 31.0 and 38.5 msec., respectively. The presentation of the auditory click increased the proportion of reports of one sensation to as high as .70. The proportion of reports of one tactile sensation is seen in Fig. 3 to be related clearly to the time of onset of the auditory stimulus. The Δt values represent in milliseconds the time by which the auditory click preceded the first of the two tactile stimuli. The highest frequency of unitary sensations occurred when the auditory click followed the second tactile stimulus by a few milliseconds for both the 31.0- and 38.5-msec. tactile intervals. Correcting for differences in sensation latency between the auditory and tactile systems would probably result in a Δt function which would indicate that maximum inhibition occurs when the auditory sensation and the second tactile sensation are simultaneous. Inhibition rapidly decreased as a function of increasing Δt when the auditory stimulus followed the second tactile stimulus (negative Δt values). Inhibition also decreased as the click was made to precede the second tactile stimulus by increasing time intervals (positive Δt values). In this case, however, the Δt functions showed a striking tendency to flatten out at Δt values near zero, i.e., when the click was simultaneous with the first tactile stimulus.

The shape of the Δt curves suggests an explanation of the manner in which an intense auditory stimulus could interfere with the temporal resolution of double tactile stimuli. Two tactile stimuli, separated by a time interval long enough so that they are perceived as successive, become difficult to perceive as two successive stimuli if the second stimulus is made less intense than the first. A much smaller reduction in temporal acuity occurs, however, when the first stimulus is made less intense than the second (Gescheider, 1966). Auditory stimulation may inhibit tactile sensations not only by elevating their absolute thresholds but also by decreasing the intensity of sensations produced by tactile stimuli above absolute threshold. Therefore, when the click and the second tactile stimulus occur together in time the apparent intensity of the second tactile stimulus would be maximally reduced, causing the proportion of "one" responses to be very high. Moreover, when the click and the first tactile stimulus were simultaneous the reduced apparent intensity of the first stimulus would cause a much smaller increase in the proportion of "one" responses.

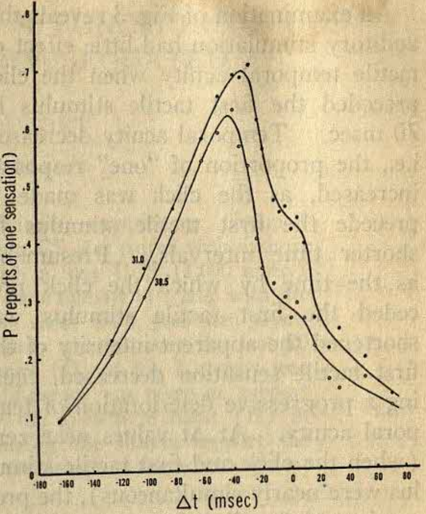


FIG. 3. Proportion of trials on which *S*s reported only one sensation plotted as a function of the amount of time by which a 90-db. SL auditory click preceded the first of two successive tactile stimuli separated by either 31.0 or 38.5 msec.

An examination of Fig. 3 reveals that auditory stimulation had little effect on tactile temporal acuity when the click preceded the first tactile stimulus by 70 msec. Temporal acuity decreased, i.e., the proportion of "one" responses increased, as the click was made to precede the first tactile stimulus by shorter time intervals. Presumably as the time by which the click preceded the first tactile stimulus was shortened the apparent intensity of the first tactile sensation decreased, causing a progressive deterioration of temporal acuity. At Δt values near zero (when the click and first tactile stimulus were nearly simultaneous), the proportion of "one" responses reached a peak, probably corresponding with maximum suppression of the first tactile sensation. However, the proportion of "one" responses again began to increase rapidly as the time interval by which the click preceded the second tactile stimulus decreased. When the click and second tactile stimulus were together in time the apparent intensity of the second tactile stimulus presumably was maximally suppressed. Temporal acuity is poorest under these circumstances (Fig. 3). As the time by which the click followed the second tactile stimulus increased, progressively less suppression of the second tactile sensation could account for the corresponding improvement in temporal acuity.

CONCLUSIONS

The results of the present experiments support the conclusion that the detectability, temporal discrimination, and perhaps the apparent intensity of touch stimuli may be significantly reduced by simultaneous auditory stimulation. Likewise, intense touch stimuli may slightly reduce the detectability of simultaneously presented auditory stimuli. The possibility of intermodality inhibition should be considered in designing communications systems which make simultaneous demands on both touch and hearing. The amount of intermodality inhibition is undoubtedly a function of many factors in addition to those investigated in the present experiments. For the effective use of a communication system it would be desirable to know how information loss caused by intermodality inhibition is affected by such factors as the frequency composition, durations, temporal relations, intensity relations, and intensity levels of the auditory and tactile stimuli.

REFERENCES

Békésy, G. v. Similarities between hearing and skin sensations. *Psychol. Rev.*, 1959, **66**, 1–22.

Elliott, L. L. Backward masking: Monotic and dichotic conditions. *J. Acoust. Soc. Amer.*, 1962, **34**, 1108–1115.

Gescheider, G. A. The resolving of successive clicks by the ears and skin. *J. exp. Psychol.*, 1966, **71**, 378–381.

Hernández-Peón, R. Reticular mechanisms of sensory control. In W. A. Rosenblith (Ed.), *Sensory communication*. Boston: Massachusetts Institute of Technology Press, 1961. Pp. 497–520.

Raab, D. H. Forward and backward masking between acoustic clicks. *J. Acoust. Soc. Amer.*, 1961, **33**, 137–139.

Schmid, E. Temporal aspects of cutaneous interaction with two-point electrical stimulation. *J. exp. Psychol.*, 1961, **61**, 400–409.

Sherrick, C. E., Jr. Effects of double simultaneous stimulation of the skin. *Amer. J. Psychol.*, 1964, **77**, 42–53.

(Early publication received
November 21, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 321–323

EYELID CONDITIONING AS A FUNCTION OF UCS DURATION WITH DRIVE EQUATED[1]

ROBERT H. DUFORT

Wake Forest University

2 groups were run in an eyelid conditioning experiment. Group 1 received 60 conditioning trials on which a CS was paired with a short UCS (50 msec.) and 60 trials on which a long UCS (1,000 msec.) was presented alone. Group 2 had the same number of trials, with the CS paired with the long UCS and the short UCS presented alone. Performance of Group 1 was significantly greater over the last 20 conditioning trials. Since the drive level (D) of the 2 groups was presumably equated, the performance difference was taken as support for Spence's reinforcement theory for the growth of habit (H) in classical aversive conditioning; specifically, the results support the drive-stimulus reduction hypothesis of the nature of the reinforcement mechanism.

Spence (1956) has developed a theory of classical aversive conditioning in which learning is governed by a reinforcement process. Specifically, the development of associative strength (*H*) is said to be a function of the intensity of the UCS. Spence prefers to limit his theorizing to this general reinforcement position rather than to specify the nature of the reinforcement mechanism (Spence, Haggard, & Ross, 1958). However, Runquist and Spence (1959) have reported an eyelid-conditioning study testing the drive-stimulus reduction hypothesis (Hull, 1943) of the mechanism of reinforcement. This hypothesis states that the offset of the aversive UCS provides reinforcement and predicts, from delay-of-reinforcement considerations, that performance in eyelid conditioning should vary inversely with the time interval between the response and the offset of the UCS. Runquist and Spence manipulated this time interval by varying the duration of the UCS. Their results gave some support to the prediction of poorer performance for groups conditioned with a long UCS than for groups conditioned with a short UCS, but the authors observe that the differences were small and of only borderline significance. They note, however, two factors in their study that might possibly operate in a direction opposite to that of UCS duration. First, the drive level (*D*) of groups conditioned with the long UCS might be greater than that of the short-UCS groups. Second, the long-UCS groups might have greater drive reduction, and hence greater habit growth, at the offset of the UCS. Both factors would tend to reduce the predicted effect of the UCS-duration variable.

An attempt may be made, using a technique developed by Spence (e.g., Homzie & Weiss, 1965; Spence et al., 1958; Trapold & Spence, 1960) for equating drive level in studies manipulating intensity of UCS, to design an experiment that controls for differences in drive level of groups conditioned with different durations of UCS. This technique consists of giving *S* an irregularly alternating order of conditioning trials, on which a CS is paired with a UCS of one intensity, and nonconditioning trials on which a UCS of another intensity is presented alone. In this way average UCS strength, and thus drive level, can be equated for two groups, one of which receives the CS paired with a strong UCS and a weak UCS alone, while the other group receives the reverse arrangement. As applied to the present problem involving UCS duration, the alternating-trials technique for equating drive would consist of giving one group conditioning trials, on which the CS is paired with a short UCS, and nonconditioning trials on which a long UCS is presented alone, while a second group is given the CS paired with the long UCS and presentations of the short UCS alone. With drive equated, such a procedure should produce a more substantial difference between the two groups than that found by Runquist and Spence, and thus provide more impressive evidence for the drive-stimulus reduction hypothesis of reinforcement in aversive conditioning.

---

[1] This research was supported by a grant from the Graduate Council of Wake Forest University. Appreciation is expressed to Susan Harbin, Howard Rollins, and Greg Olley for help in various stages of this experiment and several preliminary studies.

Method.—The *S*s were 40 male students from an introductory course in psychology. Four additional *S*s were discarded as voluntary responders (Spence & Ross, 1959). Twenty *S*s were assigned randomly to each of two groups. Following neutral instructions, *S* was given four preliminary trials, three with the CS alone and one with the UCS alone. Both groups then received 120 trials; on 60 of the trials the CS was paired with a UCS, and on 60 trials the UCS was presented alone. The trials were given in a prearranged irregular order, with the restriction that two trials of each kind occur in every block of four trials. For Group 1, the CS was paired with the short UCS (50 msec.), and the long UCS (1,000 msec.) was presented alone; for Group 2, the CS was paired with the long UCS, and the short UCS was given alone. On paired trials the CS-UCS interval was 500 msec. for both groups. The CS terminated with the UCS, giving CS durations of 550 msec. for Group 1 and 1,500 msec. for Group 2.

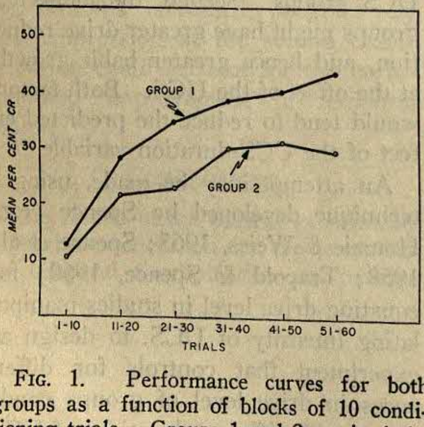


FIG. 1. Performance curves for both groups as a function of blocks of 10 conditioning trials. Groups 1 and 2 received the short UCS and the long UCS, respectively, on conditioning trials.

The apparatus was that used by Dufort and Rollins (1965). The CS was an increase in brightness from .24 to 1.08 apparent ftc. of a circular milk-glass disc. The UCS was a 2-psi air puff delivered to the right eye. A buzzer of 1-sec. duration served as a ready signal and preceded each trial by randomly varied intervals of 2, 3, or 4 sec. Intertrial intervals were 15, 20, or 25 sec., randomly arranged in advance. A CR was recorded whenever the response pen showed at least a 1 mm. deflection in the 200–500 msec. interval following CS onset.

Results and discussion.—Figure 1 presents the performance curves for both groups. The groups are essentially identical on the first 10 trials. Performance for Group 1, however, increases above that for Group 2, and Group 1 shows a clear superiority after 60 conditioning trials. The function for Group 1 appears to still be increasing while Group 2 seems to be at asymptote after 40 trials. Comparison of the curves in Fig. 1 with those of Runquist and Spence (1959) shows that a greater difference between groups was obtained in the present study. Statistical analysis shows that the difference between Groups 1 and 2 is significant over the last 20 conditioning trials, $F\ (1, 38) = 4.91$, $p < .05$, as well as over the last 10 conditioning trials, $F\ (1,\ 38) = 6.16,\ p < .025$. Similar p values are obtained with the Mann-Whitney U test.

These results are of interest in two connections. At an empirical level, they confirm and extend the findings of Runquist and Spence (1959). Specifically, they show that a substantial performance difference can be obtained between groups conditioned with short and long UCS durations if average UCS strength, and thus drive level, is equated. Following the reasoning of Runquist and Spence, the difference would be even greater if controls could be developed for the differential drive reduction in the two groups; these authors suggest, however, that it may be impossible to isolate amount of drive reduction from duration of UCS. At a more general level, the results are another demonstration that performance in eyelid conditioning, for groups with drive equated, is a function of the properties of the UCS. While studies of conditioning by other response systems (e.g., Bitterman, Reed, & Krauskopf, 1952; Mowrer & Solomon, 1954; Overmier, 1966; Wegner & Zeaman, 1958) typically report either no effect of UCS duration or better conditioning with a long duration, results such as those presented here have been taken by Spence and his co-workers to mean that associative strength in aversive conditioning follows a reinforcement principle. The present results are consistent with predictions from a particular version, the drive-stimulus reduction hypothesis, of the nature of this reinforcement principle.

REFERENCES

BITTERMAN, M. E., REED, P., & KRAUSKOPF, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 1952, **65**, 256–262.

DUFORT, R. H., & ROLLINS, H. A. Acquisition of the conditioned eyelid response under different ready-signal conditions. *Psychon. Sci.*, 1965, **3**, 81–82.

HOMZIE, M. J., & WEISS, G. UCS intensity and the associative strength of the eyelid CR with a masked conditioning procedure. *J. exp. Psychol.*, 1965, **69**, 101–103.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

MOWRER, O. H., & SOLOMON, L. N. Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *Amer. J. Psychol.*, 1954, **67**, 15–25.

OVERMIER, J. B. Instrumental and cardiac indices of Pavlovian fear conditioning as a function of US duration. *J. comp. physiol. Psychol.*, 1966, **62**, 15–20.

RUNQUIST, W. N., & SPENCE, K. W. Performance in eyelid conditioning as a function of UCS duration. *J. exp. Psychol.*, 1959, **57**, 249–252.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

TRAPOLD, M. A., & SPENCE, K. W. Performance changes in eyelid conditioning as related to the motivational and reinforcing properties of the UCS. *J. exp. Psychol.*, 1960, **59**, 209–213.

WEGNER, N., & ZEAMAN, D. Strength of cardiac conditioned responses with varying unconditioned stimulus durations. *Psychol. Rev.*, 1958, **65**, 238–241.

(Early publication received
February 23, 1967)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 324–332

EFFECTS OF FAMILIARITY ON STIMULUS RECOGNITION AND CATEGORIZATION[1]

EDWARD E. SMITH[2]

University of Michigan

In 2 different tasks Ss were instructed to press a "Yes" button if the stimulus word was a member of a predefined set of target words, and a "No" button otherwise; target sets were defined by enumeration in the E task, and by category membership in the C task. The size of the target set was either 1, 2, or 4 words or categories, and stimulus words varied in familiarity. Familiarity facilitated response times (RTs) in both tasks. The functions relating RT to size of target set in the E task were shown to reflect a memorial comparison process that follows recognition, and it was this memory process that was facilitated by familiarity.

The procedure commonly used to assess the effects of familiarity on the recognition of verbal items consists of comparing the mean tachistoscopic recognition thresholds of sets of familiar and unfamiliar items. Such a procedure: (*a*) confounds stimulus and response frequency (Goldiamond & Hawkins, 1958), and (*b*) fails to measure or control the interval between stimulus presentation and response, an interval during which much of the recognition process may be occurring. With regard to the first criticism, Goldstein and Ratleff (1961), Portnoy, Portnoy, and Salzinger (1964) and Smith and Egeth (1966) attempted to unconfound stimulus and response frequency by requiring the same response to high- and low-frequency nonsense syllables in a forced-choice task; however, the last study suggests that in such tasks *S*s adopt strategies of looking only at the initial letters of the stimuli, and thus the obtained results cannot be attributed to the familiarity of the stimuli. With regard to the second criticism, the information processing in a tachistoscopic recognition experiment may be divided into two stages: first *S* extracts information from the stimulus flash and forms a representation of it (the read-in stage); he then rapidly searches through memory for information about the possible alternatives and attempts to find a match between stimulus representation and memorial information (the memorial comparison stage). The time to read in stimulus information is only partially under *E*'s control since the visual trace persists after stimulus exposure has terminated (e.g., Mackworth, 1962), and the memorial comparisons probably take place mainly in the interval between stimulus offset and response.

---

[1] This report is based on a dissertation submitted in partial fulfillment of the requirements for the Doctoral degree at the University of Michigan. The guidance and assistance of the late P. M. Fitts is gratefully acknowledged as are the contributions of the other committee members: A. W. Melton, Thesis chairman, R. W. Pew, J. E. K. Smith, R. S. Tikofsky, and R. B. Zajonc. This research was supported in part by the Air Force Office of Scientific Research under Contract No. AF 49(638)-449 and by the Advanced Research Projects Agency of the Department of Defense under Contract No. AF 49(638)-1235 with the Human Performance Center, Department of Psychology, which was monitored by the Air Force Office of Scientific Research.

[2] Now at Behavioral and Clinical Studies Research Center, Saint Elizabeths Hospital. This research was conducted while the author was on a National Aeronautics and Space Administration Traineeship.

A procedure is needed which is capable of measuring separately both stages of recognition, and of minimizing the occurrence of response biases, and letter searching strategies. Sternberg (1963) has developed a task, which, when modified, seems to have these characteristics. On each trial a target set of elements is enumerated to *S* who is instructed to make a "Yes" response if the to-be-presented stimulus is a member of this target set, and a "No" response otherwise. When the number of target elements was varied, Sternberg (1963) found that the function relating this variable to mean response time (RT) for either "Yes" or "No" responses was linear. Sternberg (1963) proposes that when deciding whether or not a stimulus is a member of the target set, stimulus information is extracted and a representation of it formed; then this representation is compared sequentially to memorial representations of the targets. If this decision process described by Sternberg is the recognition process then the intercept of the obtained function reflects the read-in stage of recognition (as well as the time needed to execute the response), and the slope of this function reflects the rate at which the memorial comparisons are made in the second stage of recognition.

The present experiment attempts to determine where in the recognition process familiarity has its effect by ascertaining whether the intercept and/or the slope of this RT function is affected by stimulus familiarity. Since recent data indicate that the RT function may be negatively accelerated rather than linear (Chase & Posner, 1965), localization of any familiarity effect will be based on the presence or absence of an interaction of this variable with the number of target words. Such an inferred localization of a variable's effect rests on the assumptions that the duration of the read-in stage is unaffected by the number of targets, and that variations in the obtained RTs resulting from a variation in the number of targets reflect only that comparison process which underlies perceptual recognition.

This experiment also investigates the effects of familiarity in a task which requires comprehension of the stimulus words, i.e., utilization of stored information about the meanings of the stimuli. The "Yes"-"No" task is again used, a "Yes" response being required whenever the stimulus word is a member of any one of a predefined set of categories.

METHOD

Design and subjects.—Two types of tasks, target sets defined by enumeration (E task), or by category membership (C task), were combined factorially with three levels of size of target set, one, two, or four target words or categories. The six resulting conditions may be designated E1, E2, E4, C1, C2, and C4. In each condition two different stimulus orders were used, yielding a total of 12 subconditions. The *S*s (48 male and 48 female University of Michigan undergraduates) were assigned randomly to the subconditions, with the constraint that each subcondition include four males and four females; two of these males and two of these females were given one assignment of response buttons to fingers while the other four *S*s in a subcondition were given the reverse assignment.

Stimulus materials.—The stimuli were the same in all conditions and consisted of 150 English words, 50 serving as targets and 100 as nontargets. With only a rare exception, half of the targets (high-F targets) were rated as: (*a*) high-frequency associations to 1 of 12 category names (colors, tools, clothing, animals, vehicles, furniture, etc.) in the Cohen, Bousfield, and Whitmarsh (1957) norms, (*b*) high-frequency words in the Thorndike-Lorge (1944) norms, and (*c*) very familiar words on a 7-point familiarity scale by a group of 50 University of Michi-

TABLE 1

Average Characteristics of Stimulus Words

	High-F Words		Low-F Words	
	Targets	Nontargets	Targets	Nontargets
Cohen et al. Rating[a]	209.16		5.79	
Thorndike-Lorge Rating[b]	75.16	76.92	9.88	5.93
Familiarity Rating[c]	1.47	1.56	2.88	2.87
Value Rating[c]	2.79	2.84	3.29	3.68
Number of Letters	5.20	5.10	5.20	5.20

[a] For any one word, the total number of times that a group of undergraduate *S*s gave this word as an instance of a particular category, that category being one of the 12 used in the C task.

[b] Frequency of occurrence per million.

[c] Low numbers are associated with very familiar or highly valued words.

gan undergraduates (34 males and 16 females). The other half of the targets (low-F targets) were rated low on these three norms. Half of the nontargets (high-F nontargets) were rated as high frequency and very familiar on Norms (*b*) and (*c*), while the other half (low-F nontargets) were rated low on these two norms. In addition all words were rated by the group of 50 undergraduates with regard to their evaluative meaning on the good-bad scale of the semantic differential. Table 1 presents the average characteristics of these words. The correlation between the Thorndike-Lorge frequency count and the familiarity ratings for these words was −.75 while the correlation between rated familiarity and rated value was .43.

In the E1 condition each of the 50 targets was used as a target set. Each target set remained the same for a block of approximately four trials and was not used again. Within a block half the stimuli were targets (the target appearing as a stimulus twice on the average) and half were nontargets, which were never repeated and always shared some letters with the target. Target sets for the E2 and E4 conditions were formed by combining, respectively, two or four of the target sets of the E1 condition. The stimuli used in conjunction with these compound target sets were the stimuli used with the component target sets in the E1 condition but in a new randomized order. Members of a compound target set were either all high F or all low F (this constraint resulting in two targets and three nontargets being excluded from the E2 and E4 conditions), and the characteristics of nontargets (see Table 1) were independent of whether the target set they were used in conjunction with was high or low F.

In the C1 condition 12 category names (colors, tools, etc.) were used to define 12 target sets, each set consisting of all words which are exemplars of that category. Each target set remained the same for a block of 17 trials, on the average, and was not used again. Within a block half the stimuli were targets (exemplars which appeared twice on the average), and half were nontargets which were never repeated and always shared some letters with one of the targets. The number of target stimuli actually used with any category was either three, four, five, or six and all categories contained at least one high-F and one low-F target stimulus. Compound target sets for the C2 and C4 conditions as well as the stimuli used in conjunction with them were formed by the same procedure used for the E2 and E4 conditions.

Apparatus and procedure.—Stimulus words, as well as the instructions which defined the target sets, were typed (stimuli in upper-case black letters, target sets in lower-case red letters) on Esterline Angus record paper and presented in a $4\frac{5}{16}$ in. × $\frac{7}{16}$ in. window of a Gerbrands tachistoscope at a viewing distance of 23 in. At the start of each trial the target set was presented and *S* was permitted to study it for as long as he wanted; he then said "Ready" and *E* immediately depressed a foot switch which resulted, 50 msec. later, in the presentation of a stimulus and the starting of a printing timer. The *S* had previously been instructed to push a button labeled "Yes" if the stimulus was a member of the target set, and a button labeled "No" otherwise. Depression of either of the buttons, which were located in front of *S* at table height, removed the stimulus from view. Response latencies were recorded to the nearest millisecond. The *E* informed *S* whenever an error was made and the next trial began 1,500 msec. after a response. The interresponse interval, which was never less than about 3.1 sec., varied since *S*s studied larger target sets longer than smaller ones. Twenty-five practice trials preceded the 200 test trials.

The *S*s in the C conditions were instructed to inform *E* whenever they did not know the meaning of a stimulus or felt that it was ambiguous with respect to their decision. The RTs to these words (a total of 2.7% of the RTs obtained in the C conditions) were discarded. The data of five *S*s in the

C conditions who failed to know the meanings of more than 10% of the stimuli were discarded and five new *S*s were run in their places.

RESULTS

The dependence of variance, σ_i^2 on mean, $\bar{x}_i$, observed in the latency data ($\sigma \simeq \sqrt{\bar{x}_i}$) was corrected by taking square roots of the individual latencies; these transformed scores were used in all statistical analysis. In all of the analyses of variance performed, sex, finger-response button assignment, and list order were included as factors so that the variance accounted for by them would not be added to the error terms; since these variables have little relevance to the major concerns of this study they are not discussed in the following.

Enumeration task.—Figure 1 presents the mean RTs for correct "Yes" responses as a function of number of targets with stimulus familiarity and repetition (i.e., first vs. second occurrence) as the parameters.[8] High-F words were responded to faster than low-F words on both the first and second occurrences of these words, and repetition within the experimental session did not alter the magnitude of this effect. Moreover, for both first and second occurrences this effect is reflected only in the vertical displacements of the RT functions. An analysis of variance was performed on these RTs and all of the above statements were supported statistically since familiarity was a significant source of variance, $F\ (1, 24) = 15.39$, $p < .001$, while the interactions between (*a*) familiarity and repetition, (*b*) familiarity and number of targets, and (*c*) familiarity, repetition, and number of targets, were not significant. Repetition also significantly influenced RTs, $F\ (1, 24) = 25.83$, $p < .001$. Even though the second occurrence of a stimulus appeared later in the list than its first occurrence this repetition effect cannot be attributed to general practice in the task since practice (first vs. second half of experiment) was included as a factor in the analysis and although it significantly reduced RTs, $F\ (1, 24) = 48.39$, $p < .001$, it did not interact with number of targets while repetition did, $F\ (2, 24) = 7.50$, $p < .01$.

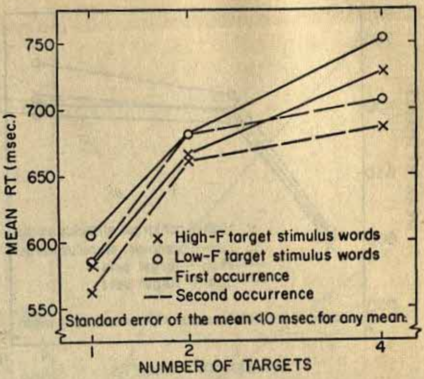


FIG. 1. Mean RTs for "Yes" responses in Enumeration conditions as a function of number of targets, with stimulus familiarity and repetition as parameters.

Figure 2 presents the mean RTs for correct "No" responses as a function of number of targets with stimulus familiarity and target set familiarity as the parameters. An analysis of variance performed on these RTs indicated that nontarget stimuli were responded to faster when *S* was looking for high-F targets than when he was looking for low-F targets, $F\ (1, 24) = 9.89$, $p < .01$, while the familiarity of the nontarget stimulus failed to reach an ac-

---

[8] The reader may recall that five stimuli from the E1 list were excluded from the E2 and E4 lists. When the RTs associated with these five words are excluded from the results for the E1 condition, the mean RTs of the untransformed scores are never changed by more than 1 msec. Hence all RTs obtained in the E1 condition were used in the graphic and statistical analysis.

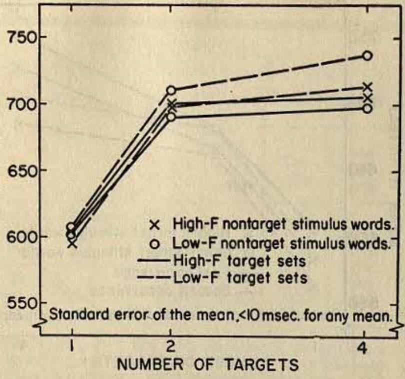


FIG. 2. Mean RTs for "No" responses in Enumeration conditions as a function of number of targets, with stimulus familiarity and target set familiarity as parameters.

ceptable level of significance, F (1, 24) = 3.95, $.05 < p < .1$. However, stimulus familiarity interacted with target set familiarity, F (1, 24) = 9.13, $p < .01$, since nontargets were responded to faster when they differed from the relevant target set with respect to familiarity level than when they were the same. An examination of Fig. 2 indicates that this interaction obtains only for low-F nontargets. The familiarity of the target set also interacted with the number of targets, F (2, 24) = 4.55, $p < .05$; again this interaction seems to obtain only for low-F nontargets (see Fig. 2).

A comparison of Fig. 1 and 2 indicates that correct "Yes" responses were faster than correct "No" responses for the one- and two-target conditions but slower than correct "No" responses for the four-target condition. To determine whether correct "Yes" responses were in general faster than *comparable* correct "No" responses the results of the above three conditions were combined, and the latencies of correct responses to first occurrences of target stimuli were compared to the latencies of correct responses to nontarget stimuli which had the same familiarity level as the relevant target set. Although the resulting t ratio did not reach an acceptable level of significance, t (47) = 1.86, $.05 < p < .1$, differential error rates (i.e., most incorrect responses were incorrect "No" responses) must be taken into consideration in interpreting these results.

The percentages of responses that were in error in the E1, E2, and E4 conditions were 2.8, 2.4, and 2.5, respectively. For every other comparison (i.e., low- vs. high-F target sets, low- vs. high-F stimuli, first vs. second occurrence, first vs. second half of experiment and "No" vs. "Yes" response) error frequency was positively correlated with mean latency.

Category task.—In Fig. 3 the mean RTs for correct "Yes" responses are presented as a function of number of categories with stimulus familiarity and

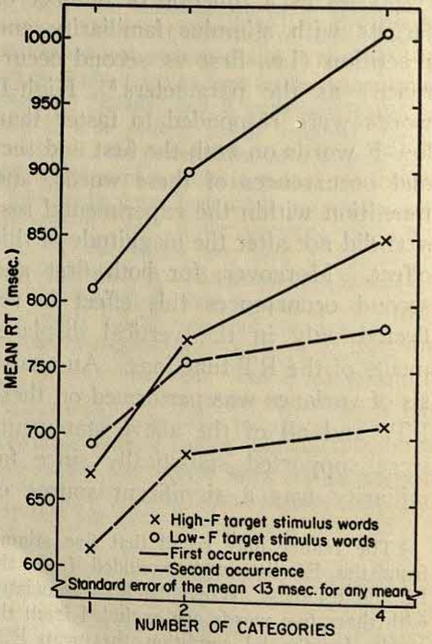


FIG. 3. Mean RTs for "Yes" responses in Category conditions as a function of number of categories, with stimulus familiarity and repetition as parameters.

repetition as the parameters. Figure 3 indicates that high-F words were responded to faster than low-F words on both the first[4] and second occurrences of the words, and on both occasions the effect appears only in the vertical displacement of the RT functions. However, this effect was substantially reduced after just one occurrence of a target stimulus in the experimental situation. These statements find statistical support when these latencies are subjected to an analysis of variance, as familiarity and the interaction between familiarity and repetition produced significant effects, $F\ (1, 24) = 409.58$, $p < .001$ and $F\ (1, 24) = 24.88$, $p < .001$, respectively, while the interactions between (*a*) familiarity and number of categories and (*b*) familiarity, number of categories, and repetition, were not significant. Repetition also significantly reduced RTs, $F\ (1, 24) = 285.25$, $p < .001$. Again the repetition effect cannot be attributed to general practice in the task (which significantly reduced RTs, $F\ (1, 24) = 7.44$, $p < .05$) since the proportion of the variance accounted for by the repetition factor was approximately 20 times greater than that accounted for by the practice factor.

In Fig. 4 the mean RTs for correct "No" responses are presented as a function of number of categories with stimulus familiarity as the parameter, and these two RT functions show the same deviations from linearity (i.e., steeper slopes between one and two categories than between two and four categories) which characterized the 12 functions already discussed. An analysis of variance performed on these RTs indicated that high-F nontarget stimuli were responded to faster than low-F nontargets, $F\ (1, 24) = 62.22$, $p < .001$, and that stimulus familiarity interacted with number of categories, $F\ (1, 24) = 3.97$, $p < .05$.

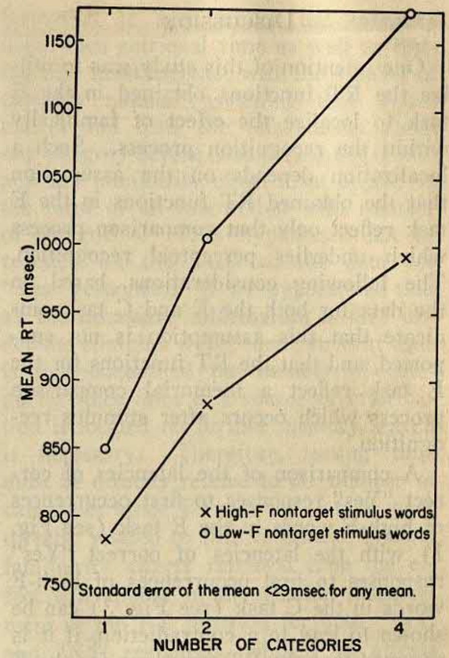


FIG. 4. Mean RTs for "No" responses in Category conditions as a function of number of categories, with stimulus familiarity as the parameter.

A comparison of the two functions in Fig. 4 with the upper two functions in Fig. 3 shows that correct "No" responses took, on the average, about 100 msec. longer than correct "Yes" responses. Error frequency and mean latency were positively correlated for every relevant comparison (i.e., C1 vs. C2 vs. C4, low- vs. high-F stimuli, first vs. second occurrence, first vs. second half of experiment and "No" vs. "Yes" response).

---

[4] The obtained familiarity effect is in no way related to differential category familiarity or difficulty since the effect can be demonstrated within a single category, i.e., the mean difference in RT between the two low-F color names and the three high-F color names was 160 msec. for first occurrences.

Discussion

One intention of this study was to utilize the RT functions obtained in the E task to localize the effect of familiarity within the recognition process. Such a localization depends on the assumption that the obtained RT functions in the E task reflect only that comparison process which underlies perceptual recognition. The following considerations, based on the data for both the E and C tasks, indicate that this assumption is not supported, and that the RT functions for the E task reflect a memorial comparison process which occurs after stimulus recognition.

A comparison of the latencies of correct "Yes" responses to first occurrences of high-F words in the E task (see Fig. 1) with the latencies of correct "Yes" responses to first occurrences of high-F words in the C task (see Fig. 3) can be shown to lead to a contradiction, if it is assumed that the RT functions for the E task reflect only that comparison process which underlies recognition. Specifically, the mean RT for this C1 condition (669 msec.) is approximately the same as that for this E2 condition (666 msec.). This implies that the time needed to recognize a stimulus drawn from a set of two alternatives is equal to the time needed to recognize the same stimulus drawn from a larger set of alternatives (assuming that the number of perceptual alternatives for an *S* in the C1 condition is greater than two) plus the time needed to categorize that stimulus on the basis of its meaning. This in turn implies (*a*) that recognition time is independent of the number of alternative stimuli, and (*b*) that the categorization process occurs simultaneously with recognition. The first implication is refuted by the RT function for correct "Yes" responses to first occurrences in the E task which indicates that recognition time increases by approximately 45 msec. (one-third of the difference between the mean RT for the E4 condition and the mean RT for the E1 condition) when the set of alternative stimuli is increased by one. The second implication can be refuted by the logical consideration that categorization cannot overlap with recognition since it is impossible for *S* to determine the meaning of a word before he has recognized it.

To resolve this paradox it must be assumed that the RTs for the E task reflect perceptual recognition plus an additional memorial comparison process, and that it is the latter which is responsible for both the 45-msec. increase in mean RT concomitant with an increase of one in the size of the target set and the "Yes" or "No" judgments. This assumption can be justified by the following consideration. In studies by Pierce and Karlin (1957) and Conrad (1962) which also required speeded identifying responses to suprathreshold verbal stimuli, i.e., reading lists of either nonsense syllables or random words, the increases in mean RTs which resulted from increasing the number of alternatives by one were less than .5 msec. Such times, which are two orders of magnitude less than the comparable times obtained in the present study, seem to be more reasonable estimates of the perceptual system's sensitivity to the number of alternatives.

Sternberg (1963, 1964), Nickerson and Feehrer (1964), Chase and Posner (1965), and Nickerson (1965) have all performed experiments similar to the present one which also yield estimates of the increase in "recognition" time per unit increase in size of set. These estimates range from 35 to 65 msec. Thus these studies are also concerned with a memorial comparison process that follows recognition and it is evident that categorization of a stimulus in an RT task may involve more than just recognition. This additional memorial comparison process may operate on the names or labels of the stimuli, i.e., the stimulus is recognized or labeled and then this label is compared to labels of the targets stored in memory. Another possibility is that this additional comparison process operates on visual features rather than labels (see Sternberg, 1964).

In view of the above what can be said about the nature of the obtained familiarity effect in the E task—is it due to a facilitation of perceptual recognition or of the subsequent memory process? The different effect of familiarity on "Yes" and "No" responses indicates that familiarity affected the process that follows recognition. Specifically, for "Yes" responses familiarity affected only the vertical displacement of the RT function, while for "No" responses familiarity interacted with the number of targets. Since the nature of the familiarity effect depended on the outcome of the memorial comparison process that followed recognition, it could not be due to the preceding recognition process. This conclusion is strengthened by the results obtained for "No" responses. Here it was possible to separate the familiarity of the stimulus from the familiarity of the target item(s) which had to be maintained in a state of "readiness" in memory; the results showed that the latter produced a greater effect than the former, indicating that the familiarity of an item becomes important when that item is involved in processes other than rapid perceptual ones. The different effect of familiarity on "Yes" and "No" responses further suggests that when the target set and the stimulus are unfamiliar *S* may check the outcome of the comparison process that follows recognition; if a "Yes" response is indicated *S* checks only that comparison which yielded a positive answer, while if a "No" response is indicated *S* checks all comparisons (targets).

In order to make a response in the C task *S* had to recognize the stimulus and then, based on an interrogation of stored information about this word, decide whether or not it was a member of any of the specified categories. The results for the E task indicate that familiarity did not influence recognition, and so must have affected either the time needed to retrieve or test the memorial information. Any attempt to specify whether retrieval or testing or both were affected by familiarity must be preceded by a consideration of how the relevant information is "filed" or indexed in memory. If we hypothesize that the relevant information is filed under the category name then retrieval time as well as testing time must increase with the number of category names (assuming that *S* retrieves and tests the information about one category before considering the next), and any familiarity effect must appear in the form of an interaction with number of categories. This hypothesis can be rejected on the basis of the data for "Yes" responses. If we hypothesize that all the necessary information is filed under the stimulus word, then retrieval time is independent of the number of categories since once this source of information is located no further memory search is necessary. Therefore, testing time must be directly related to the number of categories since RTs in the C task were directly related to this variable. Thus if familiarity affects retrieval time the effect will appear in the vertical displacement of the RT function for both "Yes" and "No" responses, while if it affects testing time the effect will appear in the form of an interaction of familiarity with the number of categories. The obtained data are consistent with this hypothesis.

REFERENCES

CHASE, W. G., & POSNER, M. I. The effect of auditory and visual confusability on visual and memory search tasks. Paper read at Midwestern Psychological Association meetings, Chicago, 1965.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. H. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

CONRAD, R. Practice, familiarity and reading rate for words and nonsense syllables. *Quart. J. exp. Psychol.*, 1962, **14**, 71–76.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relation between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

GOLDSTEIN, M. J., & RATLEFF, J. Relationship between frequency of usage and ease of recognition with response bias controlled. *Percept. mot. Skills*, 1961, **13**, 171–177.

MACKWORTH, J. F. The visual image and the memory trace. *Canad. J. Psychol.,* 1962, **16,** 55–59.

NICKERSON, R. S. Response times with a memory-dependent decision task. Paper read at American Psychological Association meetings, Chicago, 1965.

NICKERSON, R. S., & FEEHRER, C. E. Stimulus categorization and response time. *Percept. mot. Skills,* 1964, **18,** 785–793.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Syst. tech. J.,* 1957, **36,** 497–516.

PORTNOY, S., PORTNOY, M., & SALZINGER, K. Perception as a function of association value with response bias controlled. *J. exp. Psychol.,* 1964, **68,** 316–320.

SMITH, E. E., & EGETH, H. The effects of association value on perceptual search. *J. exp. Psychol.,* 1966, **71,** 687–690.

STERNBERG, S. Retrieval from recent memory: Some reaction time experiments and a search theory. Paper read at Psychonomic Society Meetings, Bryn Mawr, 1963.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction time measurements. Paper read at Symposium on Models for the Perception of Speech and Visual Form, Boston, 1964.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received July 2, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 333–341

POSITION RESPONDING AND LATENCY OF CHOICE IN SIMULTANEOUS DISCRIMINATION LEARNING[1]

SALLY E. SPERLING

University of California, Riverside

In order to examine the effect of systematic position responding on the acquisition of a discriminative response based on visual cues, the performance of rats given a standard simultaneous stimulus presentation procedure was compared with that of rats given a simultaneous procedure designed to decrease position responding. There was a significant reduction in position responding in the 2nd group, but both groups reached criterion in an equivalent number of trials and at the same rate. Both groups showed significant and equal transfer in single-stimulus successive presentation of the same 2 discriminanda. The data were interpreted as consistent with continuity theory. The latencies of correct and incorrect choices at the end of simultaneous training did not predict the immediate, appropriate differences in latencies on the transfer task.

The *E*s who use simultaneous stimulus presentation methods for discrimination studies in rats often report that position responding is extremely persistent. Sometimes *S*s do not respond to their nonpreferred side for a considerable number of consecutive trials. Since the positive and negative discriminanda are usually moved from side to side on a random or semirandom schedule, *S*s can reach approximately 50% correct choice by consistently responding to the same position. Especially if noncorrection procedure is used, the persistence of position responding may be accounted for by the intermittent reinforcement schedule that therefore obtains on the total number of responses made to the preferred side.

The criterion that *E* sets for *S*'s behavior, however, cannot be reached as long as *S* persists in position responding. Whatever the specific criterion may be, it requires correct choice considerably in excess of 50%. Eventually almost all *S*s abandon position responding and begin to make responses appropriate to the stimuli, regardless of position. Following the positive discriminandum, irrespective of its position, may be accounted for by the maximization of reinforcement that thereby occurs.

A question of theoretical interest is what effect position responding has on the emergence of the discriminative response. In developing a reformulation of noncontinuity theory, Mackintosh (1965) hypothesized that *S*s who are systematically responding to position might learn less about the discriminative stimuli than *S*s who are not. In support of this hypothesis, several experiments were cited (p. 131) that used relevant stimuli each composed of two different single cues and showed a negative relationship between the amount of learning with respect to each element.

However, the stimuli controlling position responding are presumably proprioceptive, while those controlling the requisite discriminative response are exteroceptive, regardless of whether each stimulus is composed of one or more relevant elements. It has already been suggested (Sperling, 1965a, 1965b) that these two types of stimuli do not acquire control over responding in the same way or as a function of the same variables. Therefore, the double relevant cue data may not be entirely appropriate for the determination of the effect of position responding on the acquisition of a visual discriminative response. The effect might better be examined *in situ.*

---

[1] The experimental work was supported by a National Science Foundation Grant (GB-4072). The author wishes to thank Guy Bryan, Robert Fox, and Terrance Jackson for their assistance in conducting the experiment.

In addition, since continuity theory would predict that position responding should not affect the acquisition of a visual discriminative response, a test of the two positions seemed possible.

A test was provided in the present experiment by comparing the performance of rats given a standard simultaneous stimulus presentation procedure in acquisition with that of rats given a simultaneous presentation procedure designed to decrease position responding. Both groups were then tested on a transfer task using single-stimulus, successive presentation of the same two visual discriminanda. If systematic position responding does result in interference with the acquisition of control over responding by the discriminanda, the effect could be observed in between-groups comparisons of initial or transfer data, or both. A control group received only successive training.

METHOD

Subject.—A total of 56 naive, male, hooded rats was used. They were 85–110 days of age at the start of pretraining. Four other *S*s were discarded, three for failure to reach various pretraining criteria and one who became ill during the experiment.

Apparatus.—Simultaneous acquisition training was given in an enclosed, elevated, three-segment Y maze. The apparatus was in darkness at the start of a trial except for a shielded 7-w. light on the floor of the start compartment and a red filtered photobeam light directed over the terminal end of the 12 × 3 in. start platform. When *S* broke the photobeam, a light over the stimulus compartment went on, exposing the discriminanda on the two 12 × 4 in. stimulus platforms, and a timer was started. The timer was stopped by a microswitch under the start platform when *S* stepped through a door in the partition between start and stimulus compartments and across a 1-in. gap onto one of the stimulus platforms. A retrace door between the start and the stimulus compartments and a partition between the two stimulus platforms bisecting the door between start and stimulus compartments assured that *S* could only experience both stimuli together from the choice point. Latency of response, position of response, and stimulus there were recorded.

The apparatus was adapted for single-stimulus successive presentation by placing a block over half of the door between start and stimulus compartments, permitting only one stimulus platform to show from the choice point. Latency of response was recorded.

The interior of the apparatus was painted flat black. Before the stimulus compartment light was turned on, the brightness of the stimulus platforms was $< .002$ ftl.; with the stimulus compartment light on, the brightness was 2.9 ftl.

Stimuli.—Black and white 1 in. wide Mystic tape was used to cover the stimulus platforms and their edges. One platform consisted of longitudinally laid alternating stripes (S), the other of a checkerboard (C). Stimulus platforms could be moved from side to side by an arrangement of snap fasteners. A small metal food cup could be suspended from their terminal ends.

Pretraining.—All *S*s in all groups received 14 days of pretraining during which they were tamed, food deprivation schedules based on 85% of ad lib weight were established, and training was given in crossing elevated platforms of plain pine and eating the Noyes pellets to be used in the experiments. Then 3 days of training were given in the experimental apparatus with the plain pine stimulus platforms in place. During these last 3 days, an attempt was made to equalize position preferences. On the first trial of each day, food was available in cups at the end of each pine stimulus platform. No food was present on Trials 2, 3, and 4; *S*'s previous three choices were randomly counterbalanced on Trials 5, 6, and 7 by providing food on the opposite side.

Experimental groups and procedure.—The *S*s were given 12 trials a day throughout acquisition and transfer on a 10-min. intertrial interval in squads of up to six at a time. Reinforcement consisted of two 94 mg. Noyes pellets in the food cup at the end of the positive platform. No food cup was present with the negative platform.

Nineteen *S*s received simultaneous discrimination training using a standard, semirandom schedule for stimulus position (Group NRC). Each platform appeared six times on each side within a day in random order, with the restriction that the same configuration could not be presented for more than three consecutive trials. Different schedules were used within and across squads. For 10 *S*s, S was the positive platform and C the negative; S was negative and C positive for the other 9 *S*s.

Nineteen *S*s received simultaneous discrimination training using a response contingent procedure (Group RC) adapted from Abordo and Rumbaugh (1965). On the first trial of the first day, *S*'s positive platform was on his preferred side (as judged from apparatus pretraining) and the negative platform was on his nonpreferred side. The position of the two stimuli was maintained until *S* responded to the positive. On the next trial, the position of the positive and the negative was reversed. They remained in that configuration until *S* again responded to the positive. On the next trial their positions were reversed. This procedure was continued throughout acquisition.

Noncorrection procedure with a 20-sec. detention was used for all 38 *S*s and training was continued to a criterion of not more than two incorrect choices on 1 day followed by not more than one incorrect choice on the next day—.875 correct over the 2 days.

On the day following the second criterion day, each *S* was given single-stimulus, successive presentation of the same two stimuli, maintaining their positive and negative values. There were six positive and six negative presentations a day, in random order within and across squads, with the restriction that the same stimulus could not be repeated more than three consecutive times. Training was continued to the criterion that at least five latencies of response on the six negative trials had to be longer than any of the six latencies of response on the positive trials within a 12-trial day.

The single stimuli appeared consistently on the preferred side for 10 of the *S*s in each simultaneous acquisition group and on the nonpreferred side for the other 9 *S*s in each group.

A third group of 18 *S*s was given only single-stimulus, successive presentation of the two discriminanda to the same latency criterion (Group SuC). Within this group, positive and negative stimuli were counterbalanced as was whether the single stimuli appeared on the preferred or nonpreferred side (side preferences were determined from apparatus pretraining).

A white noise generator provided a masking noise during the experimental sessions.

RESULTS

Neither of the procedural counterbalancings led to any reliable differences in performance. The data were, therefore, combined within each of these categories.

Except for the tests on number of trials to criterion, the analysis of all results for all groups and all conditions was based upon Vincentized sixths of trials to criterion for the appropriate data.

Simultaneous acquisition.—Figure 1 shows mean proportion of responses to the preferred side for each group for each sixth of acquisition. Position preferences were determined from the data; a random assignment was made for one *S* in each group not showing a clear preference.

An analysis of variance was computed on the individual data, with sixths of acquisition as a dependent dimension. Although both main effects were significant, a significant Groups × Sixths interaction, F (5, 180) = 12.22, $p < .001$, permitted a direct test of the difference between the two groups for five of the six acquisition blocks. For each of the first five sixths, Group NRC showed significantly more position responding than Group RC, F (1, 180) = 14.37, 12.20, 77.69, 55.80, and 5.32, respectively; $p < .001$ for the first four tests and $p < .025$ for the fifth. The amount of position respond-

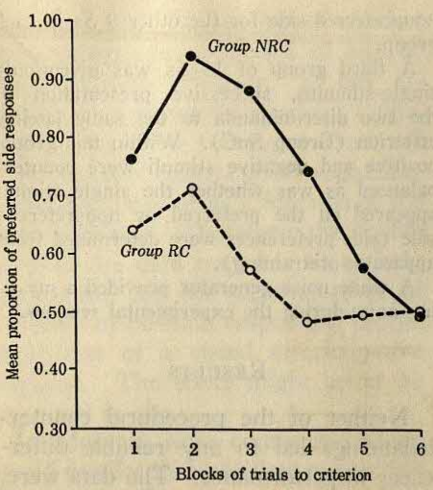


FIG. 1. Mean proportion of total responses that were made to the preferred side by sixths of trials to criterion for Group NRC and Group RC.

ing, therefore, was both immediately and persistently less in Group RC than in Group NRC.

However, although the response contingent procedure did lead to decreased position responding in Group RC, the reduction was not reflected in any difference in either trials to criterion or in the pattern of correct responses to criterion. Table 1 shows mean trials to criterion for the various groups for the different parts of the experiment. The difference in simultaneous acquisition between 84.0 trials for Group NRC and 77.7 trials for Group RC was not reliable, t $(36) = 0.82$; $p > .40$. The standard deviations of 24.41 and 24.45, respectively, were essentially identical.

TABLE 1

MEAN TRIALS TO CRITERION FOR INITIAL AND TRANSFER TASKS

Groups	Mean Trials to Criterion	
	Simultaneous	Successive
Simultaneous Training-Successive Transfer		
Group NRC ($N=19$)	84.0	16.4*
Group RC ($N=19$)	77.7	22.1*
Successive Training		
Group SuC ($N=18$)	—	76.0

* $p < .01$ for test against Group SuC mean.

Rank-order correlation coefficients computed within each group indicated that there was no relationship between strength of position responding and trials to criterion in simultaneous acquisition for either Group NRC or Group RC.

During acquisition, *S*s sometimes responded correctly and sometimes incorrectly. Looking only at the correct responses, some were made to the positive when it was on *S*'s preferred side and some when it was on *S*'s nonpreferred side. Because of position responding, some *S*s made no responses at all to the nonpreferred side during some sixths. In addition, some *S*s made no correct responses at all during some sixths. The problem of missing data necessitated an analysis based on group data rather than on individuals within a group. Ratios were, therefore, computed separately for each group and for each sixth for the proportion of the total responses made to each side that were correct responses. These group mean proportions were used in a Groups × Position × Sixths of Acquisition analysis of variance. A significant main effect for Sixths, F $(5, 5) = 12.57$; $p < .01$, indicated that the accuracy of performance increased with training. Since neither of the other two main effects approached significance, the rate of approach to criterion appeared to be the same for both groups and for preferred and nonpreferred sides. Therefore, the differing frequency with which responses were made to preferred and nonpreferred sides apparently was not associated with any differing probability of correct choice on the two sides. These data, collapsed over position, are shown in Fig. 2.

The latencies with which responses were made in simultaneous acquisition were transformed by logarithm (latency × 10). As already discussed, not all *S*s responded to both the positive and negative stimuli on both the preferred and nonpreferred sides during each sixth of acquisition. This resulted in there being some empty cells for some *S*s when mean log latency of response during each sixth was computed separately for each stimulus and each side. In order to analyze the latency data, group means were again computed. For a given sixth and side, none of an *S*'s latencies of response were used if there were not responses to both the positive and the negative discriminanda on that side in that sixth. If *S* had responded at least once to both the positive and the negative on

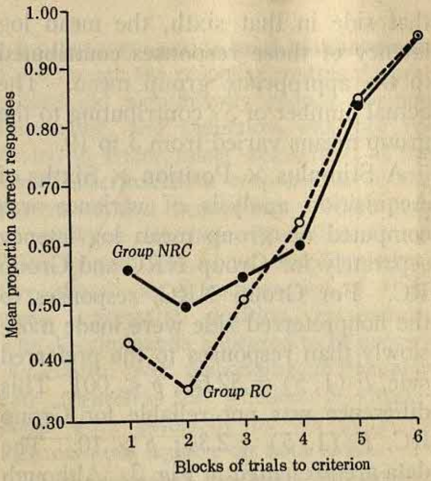


FIG. 2. Mean proportion correct responses by sixths of trials to criterion for Group NRC and Group RC.

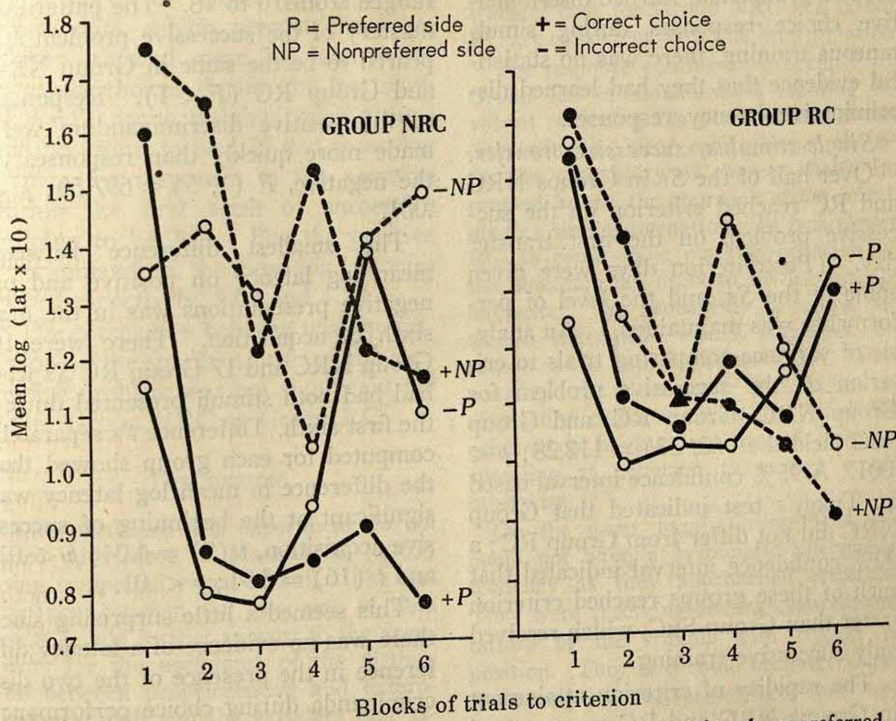


FIG. 3. Mean latency of correct and incorrect choices on preferred and nonpreferred sides by sixths of trials to criterion for Group NRC and Group RC.

that side in that sixth, the mean log latency of those responses contributed to the appropriate group mean. The actual number of *S*s contributing to the group means varied from 3 to 19.

A Stimulus × Position × Sixths of Acquisition analysis of variance was computed on group mean log latency separately for Group NRC and Group RC. For Group NRC, responses to the nonpreferred side were made more slowly than responses to the preferred side, $F\ (1, 5) = 57.66$; $p < .001$. This difference was not reliable for Group RC, $F\ (1, 5) = 2.31$; $p > .10$. The data are presented in Fig. 3. Although each group came to respond faster on correct than on incorrect choices, regardless of the position of the discriminanda, the difference in mean log latency was not significant within or over sixths for either group. Even though the two groups had learned discriminative choice responses during simultaneous training, there was no statistical evidence that they had learned discriminative latency responses.

Single-stimulus, successive transfer.—Over half of the *S*s in Groups NRC and RC reached criterion on the successive problem on the first transfer day. (Postcriterion days were given some of the *S*s, and the level of performance was maintained.) An analysis of variance comparing trials to criterion on the successive problem for Group NRC, Group RC, and Group SuC yielded $F\ (2, 53) = 112.28$; $p < .001$. A 95% confidence interval based on Tukey's test indicated that Group NRC did not differ from Group RC; a 99% confidence interval indicated that each of these groups reached criterion faster than Group SuC, which received only successive training.

The rapidity of criterion attainment in Groups NRC and RC as compared to Group SuC seemed to be due to the fact that the first two groups showed the same immediate, appropriate difference between the latencies of responses to the positive and negative discriminanda.

An analysis of variance comparing the two groups was computed for mean log latency for responses to the positive and the negative by sixths of successive acquisition. Group data were used for the analysis, with the data from only those *S*s who had had both positive and negative presentations in a given sixth contributing to the group means for that sixth. (Since an average of 3.2 trials constituted each sixth and since stimulus presentation was semirandom, not all *S*s had both positive and negative presentations during each sixth even though all *S*s responded on all trials.) The actual number of *S*s contributing to the various means ranged from 10 to 18. The pattern of mastery of the successive problem appeared to be the same in Group NRC and Group RC ($F < 1$). Responses to the positive discriminandum were made more quickly than responses to the negative, $F\ (1, 5) = 697.50$; $p < .001$.

The smallest difference between mean log latency on positive and on negative presentations was in the first sixth of acquisition. There were 10 Group NRC and 17 Group RC *S*s who had had both stimuli presented during the first sixth. Difference *t*'s separately computed for each group showed that the difference in mean log latency was significant at the beginning of successive acquisition, $t\ (9) = 4.04$; $p < .01$, and $t\ (16) = 3.91$; $p < .01$.

This seemed a little surprising since there was no evidence of a latency difference in the presence of the two discriminanda during choice performance on the simultaneous problem. Because the main effect for Stimuli was not significant in the analysis of the log latency data for correct and incorrect choices, no legitimate test of the difference at the end of simultaneous acquisition could be made. However, there were six Group NRC and eight Group RC *S*s who had made at least one error during their last sixth of acquisition, giving them mean log latency scores for both correct and incorrect choices, and who also had had presentations of both the positive and negative stimuli during the first sixth of successive acquisition. Difference *t*'s indicated an abrupt increase in mean log latency of response to the negative compared to the mean log latency with which the negative was chosen, $t\ (5) = 4.48$; $p < .01$, and $t\ (7) = 2.46$; $p < .05$. The comparable data for responses to the positive and choices of the positive did not approach significance for either group. The group data, which present the same picture, showed mean log latency for incorrect choices during the last sixth of simultaneous training, over groups and positions, to be 1.23 and mean log latency for responses to the negative during the first sixth of successive training to be 2.11. For the positive discriminandum, the mean log latencies were, respectively, 1.02 and 1.28.

The correlation between trials to criterion on the simultaneous and the successive problem was not significant for either Group NRC or Group RC.

DISCUSSION

The results of this experiment do not agree with those of Mackintosh (1965) using compound visual stimuli. The discrepancy between the two sets of data seems to add further support to the arguments for the usefulness of the distinction between proprioceptive and exteroceptive stimuli and the variables that affect their control over responding.

A continuity interpretation of the simultaneous acquisition data of the present experiment seems most parsimonious. Even though the response contingent procedure did significantly reduce position responding, the reduction was not reflected in either faster acquisition of the visual discriminative response or in the pattern of that acquisition. There was no evidence in these data that Group NRC learned less about the discriminative stimuli as a consequence of their more systematic position responding. The discriminanda appeared to be acquiring control over responding at the same rate regardless of whether *S* experienced them primarily on his preferred side or more equally on both sides.

Two additional findings add further support to the above statements. First, the lack of any within-groups correlation between the strength of position responding and trials to criterion during simultaneous acquisition also suggests that position responding does not affect visual discrimination learning. Second, the transfer data for Group NRC and Group RC are consistent with this interpretation.

The response contingent procedure for Group RC in simultaneous training involved reversing the position of the two discriminanda after each correct choice. As the probability of correct choice increased, so did the approach of the schedule to a single alternation. In fact, Group RC *S*s could have reached criterion on the simultaneous problem by learning to alternate. The probability of simple response alternation would, therefore, increase as *S*s approached criterion. Presumably the increased probability of alternation could result in the visual discriminanda having less control over responding at criterion than they would otherwise.

On the other hand, Group NRC *S*s could not have a systematic alternation tendency as they approached criterion. They were receiving semirandom presentations of the stimuli with respect to position. They also could no longer have the strong position preferences that they exhibited early in training, and they actually were responding randomly with respect to position at criterion. Presumably, then, the visual discriminanda would have primary, perhaps exclusive, control over responding by the end of simultaneous acquisition.

Any difference between these two groups in the degree of the control over responding by the discriminanda might be expected to appear in differences in the mastery of the single-stimulus, successive transfer task even though there was no difference in initial acquisition. Yet these two groups did not differ from each other in either gross trials to criterion or in the pattern of the latencies to criterion. Furthermore, the transfer data demonstrated that the control over responding acquired by the discriminanda during simultaneous acquisition in each group not only was equal but also was considerable. Both of these groups reached criterion on the transfer task with over 70% savings compared to *S*s who received only the single-stimulus, successive task.

The transfer data for Group NRC and Group RC are even more impressive when it is remembered that a significant difference in the latencies of response on positive and on negative presentations was required in the single-stimulus, successive task. Although the difference in mean log latency of correct and incorrect choices was in the right direction at the end of simultaneous training, there was no statistical evidence that the difference was reliable.

It seems unlikely that an appeal to the acquired distinctiveness of the visual discriminanda as a consequence of simultaneous training (Lawrence, 1949, 1950) will account for the successive transfer data. Acquired distinctiveness should make it easier to learn a new response to these same stimuli. However, many of the *S*s in this experiment apparently did not have to learn the latency difference in response at all; it was already there at the beginning of successive transfer, even though it was not observable at the end of simultaneous training.

The latency measure used in simultaneous acquisition was a fairly good choice time measure. It seemed reasonable to expect that there might come to be a difference in the latencies with which *S*s made correct as opposed to incorrect choices, perhaps as a consequence of the different delays of reinforcement that followed each. Indeed, Mahut (1954) reported such a difference for responding to the preferred side of the 8 of 39 *S*s for whom choice time measures were taken (latencies for nonpreferred side responding were not reported). Mackintosh (1965) mentioned that he had observed the same phenomenon (p. 126).

It is possible that in the present study a real difference in the latencies of correct and incorrect choices did exist by the end of initial acquisition even though it was not statistically demonstrable. However, even if this were the case, the rapidity with which Groups NRC and RC reached criterion on the successive transfer task would not be predictable from the simultaneous acquisition latency data.

Another analysis of the data also assumes that differential latencies of response to the positive and the negative stimuli may have been acquired during simultaneous acquisition but that the nature of a choice error, at least late in acquisition, precludes the observation of the real latency of choice of the negative stimulus. If the choice of or approach to the negative instead of the positive is regarded as being due to an error in classification of the negative, with *S* literally mistaking the negative for the positive, then an incorrect choice would be made with the latency appropriate to a correct choice, and the two would not be expected to be different.[2] In that case, the latency of choice data could not be expected to predict the rapidity of the transfer to a task requiring a latency difference. They would reflect only the latency of response appropriate to the positive and contain no information about any acquired difference in latency of response associated with the two stimuli.

With respect to these speculations, it

---

[2] Grateful acknowledgment for this suggestion is made to John Theios, who, however, is not responsible for the use made of it here.

is perhaps suggestive that there was no difference in the latency of response on positive presentations at the beginning of successive training and the latency of correct (and, therefore, incorrect) choices at the end of simultaneous acquisition. Furthermore, it is possible that the changes in environment that accompanied the change from simultaneous acquisition to successive transfer made a classification error of the negative stimulus less likely, perhaps by making it more likely that *S* would visually examine the discriminanda in the new situation. If so, the immediately observable longer latency of response on presentations of the negative in successive training could be reflecting a response already learned during simultaneous acquisition but not observable there for the above reasons.

REFERENCES

ABORDO, E. J., & RUMBAUGH, D. M. Response-contingent learning-set training in the squirrel monkey. *Psychol. Rep.*, 1965, **16**, 797–802.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

MAHUT, H. The effects of stimulus position on visual discrimination by the rat. *Canad. J. Psychol.*, 1954, **8**, 130–138.

MACKINTOSH, N. J. Selective attention in animal discrimination learning. *Psychol. Bull.*, 1965, **64**, 124–150.

SPERLING, S. E. Reversal learning and resistance to extinction: A review of the rat literature. *Psychol. Bull.*, 1965, **63**, 281–297. (a)

SPERLING, S. E. Reversal learning and resistance to extinction: A supplementary report. *Psychol. Bull.*, 1965, **64**, 310–312. (b)

(Received July 5, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 342–350

WHAT IS THE NECESSARY AND SUFFICIENT CONDITION FOR REINFORCEMENT IN THE CONTINGENCY SITUATION?[1]

ROBERT EISENBERGER
University of California, Santa Barbara

MICHAEL KARPMAN
California State College, Los Angeles

AND JAMES TRATTNER
University of Chicago

A set of experiments was initiated to test Premack's hypothesis that any more probable response will reinforce any less probable response. In the 1st experiment, the more probable response reinforced the less probable response only when the contingent behavior was suppressed beneath the free-performance level. Hypothesizing the necessary and sufficient condition for reinforcement to be the animal's necessity of increasing instrumental responding if it is to maintain contingent responding at the free-performance level, it was correctly predicted that upon establishing such a condition the less probable response would reinforce the more probable response. The general conditions underlying such "response suppression" are denoted mathematically, and the specific conditions associated with the reinforcement effect in Premack's confirmatory experiments are derived as a special case of response suppression.

Premack (1959, 1961, 1965) hypothesized any more probable response will reinforce any less probable response, the response probabilities being assessed when the animal may freely perform either behavior ("operant period"). Predictions based on this hypothesis led to the demonstration that traditional goal behaviors could themselves be reinforced. Recording the amounts of candy eating and pinball-machine manipulation in an operant period, Premack (1959) reported that in accord with the traditional hedonistic viewpoint, some children (those with the higher operant probability of eating than manipulating) subsequently increased manipulation above operant level for the opportunity to eat and did not increase eating for the opportunity to manipulate. However, those children with the higher probability of manipulating than eating, increased eating above operant level for the opportunity to manipulate and did not increase manipulation for the opportunity to eat. Further, using a Cebus monkey, manipulatory responses were shown to reinforce other less probable manipulatory responses, but not more probable manipulatory responses (Premack, 1963). And, to show that the reinforcement relation between two behaviors could be reversed for the individual *S*, parameters were manipulated to make activity-wheel running by rats more probable than drinking and, in

---

[1] The senior author wishes to thank Seymour Levitan, Joseph Phelan, and Herbert Moskowitz for their encouragement, and Leonard Chatis for his ingenuity in design and construction of equipment.

Some of the results were presented in a paper at the Western Psychological Association Convention, Long Beach, April 1966. The first three experiments were performed at California State College, Los Angeles, the fourth experiment at University of California, Santa Barbara. Of the junior authors, Karpman aided in the planning and execution of Exp. I and Trattner in preliminary exploratory studies.

the same *S*s at another time, drinking more probable than running. In both situations, upon instituting a contingency, the more probable response reinforced the less probable response (Premack, 1962).

EXPERIMENT I

Previous experiments utilized rats, monkeys, and young children at low or moderate operant probabilities of the instrumental response (less than .15, where reported: see Premack, 1963; Schaeffer, 1965). Experiment I used young adults and tested whether probability differential (i.e., the higher free-performance probability of the contingent response than instrumental response) is sufficient for reinforcement at both high and low probabilities of the instrumental response.

Method

Subjects and apparatus.—Thirty-nine college students, most of whom were enrolled in introductory psychology courses and a few with no course work in psychology, served as *S*s. A Wahman activity wheel was equipped with a solenoid-operated brake which, when deactivated, made contact with the wheel. The brake was not visible to *S*s. A wooden handle was attached to the rim of the wheel, and a bar press was placed nearby; turning the wheel by its handle and pressing the bar were intended as manipulatory responses. A specially designed control panel allowed *E* to provide *S*s with an operant period (unlocked bar and wheel) or to set an instrumental requirement (a given number of bar presses required of *S*s to unlock the wheel, and duration of time the wheel subsequently remained unlocked). Bar presses and wheel turns were recorded on digital counters and on an Esterline Angus recorder. Duration of time spent turning the wheel was recorded on a Standard electric timer.

Procedure.—Each *S* was seated in front of the wheel and bar press, and told:

> Please sit and remain in this chair until you are told the experiment is over. Before you, you see a wheel with a wooden handle on it and also a metal bar. [Experimenter pointed to each of the items as he named it.] When I say to you "go ahead," you may, with either one or both hands, press the metal bar or turn the wheel by its handle. How you spend your time is up to you so long as you remain seated. You may press the bar; you may turn the wheel; you don't have to press the bar; you don't have to turn the wheel; or you can do both at the same time. I repeat—you may press the bar; you may turn the wheel; you don't have to press the bar; you don't have to turn the wheel; or you can do both at the same time.
>
> Do you have any questions about these directions?
>
> All right. Go ahead.

Each *S* thereupon received a 5-min. operant period, followed by a 5-min. contingency period in which the wheel could be unlocked for 10 sec. at a time by one press of the bar. The sound of the brake clearly signaled the locking and unlocking of the wheel during the contingency period. Each succeeding instrumental requirement could be met only after the wheel had relocked; keeping the wheel unlocked for the full 5 min. would have necessitated one bar press every 10 sec., for a total of 30 presses.

Results

Premack's probability differential hypothesis predicted reinforcement for *S*s having the higher free-performance probability of the contingent response (wheel turning) than instrumental response (bar pressing). Of the 39 *S*s, 36 spent more time in the operant period turning the wheel than pressing the bar, and therefore had a higher average probability of wheel turning than bar pressing. These 36 *S*s were divided into two groups, the higher-than-median operant bar pressers and the lower-than-median operant bar pressers. The "High" bar press group had a median operant probability of bar pressing of .22; the "Low" bar press group's median probability was .01. A statistically significant proportion of the Low bar pressers increased their number of bar presses from the operant to contingency period, 14 out of 18 *S*s with 1 not changing, $p < .01$, one-

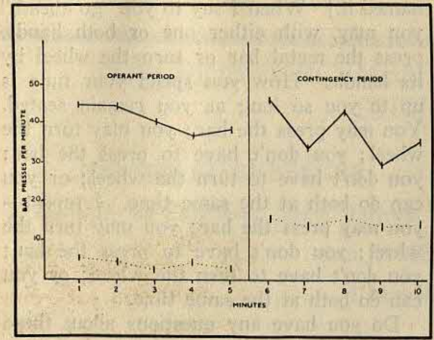


FIG. 1. Mean number of bar presses (instrumental response) performed by the High bar pressers (solid line) and by the Low bar pressers (dotted line) each minute of the operant and contingency periods. (Data was available for 15 of the 18 High bar pressers and all 18 of the Low bar pressers.)

tailed binomial sign test. However, the proportion of High bar pressers showing an increase did not exceed that expected by chance alone, 7 out of 18 *S*s, $p = .6$, one-tailed binomial sign test. The proportion of Low bar pressers surpassing their operant rate was significantly greater than that of the High bar pressers, χ^2 (1) = 8.7, $p < .005$ (Fig. 1 shows for each group the mean number of bar presses performed each minute of the operant and contingency periods). Thus the more probable response was demonstrated to reinforce the less probable response only for *S*s having a relatively low free-performance probability of the instrumental response.

Previous experiments (Premack, 1961, 1963) suggested two plausible explanations, still assuming validity of the probability differential hypothesis, for failure to detect the reinforcement effect among High bar pressers. First, the method of determining if a given *S* had a higher free-performance probability of wheel turning than bar pressing was to compare the total operant durations of the two responses. Such an approximation would not be sensitive to some reversals of response-probability order near the close of the operant period. As a test for response-probability reversals among *S*s who did not increase instrumental responding from the operant to contingency period, each such *S*'s difference in duration of time spent wheel turning and bar pressing for the first operant minute was compared with his difference for the last operant minute. The necessary data was available for 13 of the 15 *S*s who showed no increase in instrumental responding. For none of these *S*s was there a reversal of response-probability order: 8 *S*s increased their margin of wheel turning over bar pressing, 3 still maintained a difference of over 100%, 1 a difference of 87%, and 1 a difference of 3%.

A second plausible explanation was that the reinforcement cccurring for High bar pressers was masked, due to a decline throughout the experimental session of the free-performance level of the instrumental response (bar pressing). To control for such a possibility, for each *S* in the High and Low bar-press groups the change in number of bar presses from that in the first operant minute to that in the last operant minute was extrapolated to the last minute of the contingency period. The *S*s whose number of bar presses in the last minute of the contingency period exceeded their extrapolated operant rate, were counted as having been reinforced. Of the 18 Low bar pressers, 15 surpassed their extrapolated rate, 1 fell below it, and 2 equaled it, $p < .005$, of as high a proportion by chance alone, one-tailed binomial sign test. Of the 15 High bar pressers for which the necessary data was available, 7 surpassed their extrapolated rate while 8 fell below it, $p = .7$, of as high a proportion by chance alone, one-tailed binomial sign test. Further, the proportion of Low bar pressers exceeding their extrapolated rate was significantly greater than that of the High bar pressers, χ^2 (1) = 5.2, $p < .025$. Thus, use of a reinforcement criterion sensitive to changes in free-performance level of instrumental responding did not eradicate the failure to demonstrate reinforcement for *S*s having the high free-performance level of instrumental responding.

EXPERIMENT II

This failure suggested three alternative reformulations of the necessary and sufficient conditions for reinforcement:

1. *Probability differential-low instrumental probability.* In addition to the higher probability of the contingent response than instrumental response, the probability of the instrumental response must be sufficiently low.

2. *Probability differential-response suppression.* In addition to probability differential, the instrumental requirement must be sufficient to suppress contingent responding beneath its free-performance level.[2] Such "response suppression" did not occur for most of the first experiment's High bar pressers; their operant bar-press level was sufficiently high and the instrumental requirement sufficiently low that without increasing instrumental responding above operant level, the *S*s were still able to turn the wheel for as long a duration of time, on the average, in the contingency period as they had in the operant period (see Fig. 1 and 2). By definition, response suppression occurs when it is necessary for an individual to increase instrumental responding above free-performance level if he is to be able to perform the contingent response the same amount in the contingency period as he would in the free-performance situation.

3. *Response suppression.* The necessary and sufficient condition for reinforcement is response suppression. Any response serving to overcome response suppression will be reinforced.

This experiment tested between the alternative hypotheses.

Method

Subjects and apparatus.—Twenty-seven college students, enrolled in introductory psychology courses, served as *S*s. Equipment used was the same as in Exp. I.

Procedure.—Were the *response suppression* hypothesis correct, by establishing response suppression the high-probability response would be reinforced (vs. Hypothesis 1) and the less probable response would reinforce the more probable response (vs. Hypotheses 1 and 2). Wheel turning was now used as the instrumental response, with bar pressing as the contingent response. This new contingency arrangement was appropriate since judging by the first experiment, most *S*s would be expected to have

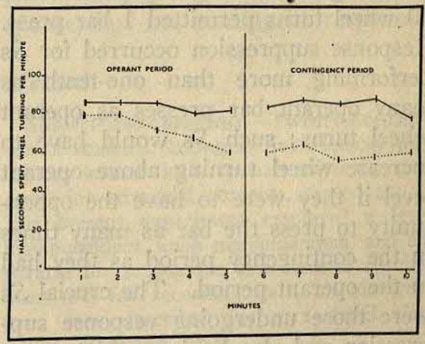


FIG. 2. Mean duration of time spent turning the wheel (contingent response) by the High bar pressers (solid line) and by the Low bar pressers (dotted line) each minute of the operant and contingency periods. (Data was available for 14 of the 18 High bar pressers and 15 of the 18 Low bar pressers.)

---

[2] Premack (1965), using rats, reported "response suppression" a necessary condition for reinforcement in the contingency situation. Integrating this finding with his probability differential principle would yield the *probability differential-response suppression* hypothesis discussed here.

a high operant wheel-turn probability, with wheel turning more probable than bar pressing. A 5-min. operant period was followed by a 5-min. contingency period in which a high instrumental requirement was employed (each 10 wheel revolutions afforded *S*s the opportunity to press the bar once), helping ensure suppression of the contingent response well beneath its free-performance level. The instructions to *S*s preceding the operant period remained unchanged from the first experiment, except *S*s were now instructed not to turn the wheel and press the bar at the same time. Following the operant period, each *S* was told:

> Please stop until told to proceed. From now on, you do not have to turn the wheel; you may turn the wheel whenever you wish; you do not have to press the bar; you may press the bar whenever you wish. But for each bar press you wish to make, you must first turn the wheel 10 complete revolutions. I repeat—for each bar press you wish to make, you must first turn the wheel 10 complete revolutions. You may turn the wheel without pressing the bar, but you must make at least 10 revolutions for each bar press. You may only press the bar once at a time.
>
> Do you have any questions about these directions?
>
> All right. Go ahead.

Results

During the contingency period each 10 wheel turns permitted 1 bar press. Response suppression occurred for *S*s performing more than one-tenth as many operant bar presses as operant wheel turns; such *S*s would have to increase wheel turning above operant level if they were to have the opportunity to press the bar as many times in the contingency period as they had in the operant period. The crucial *S*s were those undergoing response suppression and also having a higher operant probability of the instrumental response than contingent response. The two conditions were met by 21 of the 27 *S*s. The median operant probability of wheel turning (instrumental response) for the 21 *S*s was .35, compared to the operant probability of .22 for *High* bar pressers in the first experiment. Thus the *probability differential-low instrumental probability* hypothesis and the *probability differential-response suppression* hypothesis predicted absence of the reinforcement effect, while the *response suppression* hypothesis did predict reinforcement. Of the 21 crucial *S*s, 16 performed more wheel turns in the contingency period than operant period, $p < .005$, of as high a proportion by chance alone, one-tailed binomial sign test. The median number of 45° wheel movements rose from 323 in the operant period to 695 in the contingency period. The *response suppression* hypothesis was therefore supported.

EXPERIMENT III

Using rats, it has been reported that eliminating the opportunity to engage in a normally performed behavior may increase the daily free-performance levels of alternative behaviors (Premack & Premack, 1963); it was therefore possible that the increase in wheel turning in Exp. II stemmed merely from *S*s' reduced opportunity to press the bar during the contingency period, and not from wheel-turning's instrumental function of allowing bar pressing. As a control, this experiment tested the effect upon wheel turning of merely limiting the opportunity to press the bar.

Method

Subjects and apparatus.—Eleven college students, enrolled in introductory psychology courses, served as *S*s. Equipment was the same as in the first two experiments.

Procedure.—The *S*s received the same initial instructions and preliminary 5-min. operant period as in Exp. II. Immediately thereafter, a 5-min. "limited" operant period ensued in which *S*s were allowed to turn the wheel, but not press the bar. Each *S* was told prior to the limited operant period:

> Please stop until told to proceed. From now on, you do not have to turn the wheel; you may turn the wheel whenever you wish; but you must not press the bar. I repeat—you do not have to turn the wheel; you may turn the wheel whenever you wish; but you must not press the bar.
>
> Do you have any questions about these directions?
>
> All right. Go ahead.

Results

The crucial *S*s were those who in the first operant period performed more than one-tenth as many bar presses as wheel turns, such having been the criterion for selecting the second experiment's crucial *S*s. If the increase in instrumental responding (wheel turning) in the second experiment were not merely the result of limiting the opportunity to press the bar, the present experiment's *S*s would be expected to show a lesser increase in number of wheel turns from the initial operant to limited operant period than the previous *S*s' increase from the operant to contingency period. Sixteen of the 21 crucial *S*s in the former experiment increased their number of wheel turns, while all 8 crucial *S*s in this experiment showed decreased wheel turning. The median number of 45° wheel movements rose from 323 to 695 in Exp. II, while declining from 226 to 106 in this experiment. Thus the increase in instrumental responding in the former experiment was not attributable to the effect of merely limiting the opportunity to perform the contingent response.

EXPERIMENT IV

In Exp. II, *S*s were told of the instrumental requirement prior to the start of the contingency period (each 10 wheel turns allowing 1 bar press). It had not been possible to present *S*s with a locked bar, having them discover for themselves the instrumental requirement, because the apparatus was not designed for the turn-in-order-to-press contingency. The question arose whether instructing *S*s as to the instrumental requirement might have "induced" the increase in wheel turning. The present experiment employed new apparatus to eliminate the necessity for such instructions. As an additional reinforcement criterion, this experiment tested whether response suppression would subsequently yield a higher-than-operant rate of instrumental responding in an extinction period (period in which the instrumental response is no longer reinforced). Also, to control for the (possibly reinforcing) sound associated with the locking and unlocking of the apparatus during the contingency period, the sound was presented throughout the experimental session on the same schedule as in the contingency period (after every tenth instrumental response and after each contingent response). The extinction period replicated the operant period in that both the instrumental response and contingent response could be freely performed (vs. the more usual extinction procedure of allowing free performance of the instrumental response, but not contingent response).

Method

Subjects and apparatus.—Twenty-five high-school students, taking an introductory psychology course in a special university program for advanced students, served as *S*s. The present experiment employed a lever manipulandum, knob manipulandum, and apparatus for setting the instrumental requirement and for recording data, all of which have been described in more detail elsewhere (Premack, 1963).

Procedure.—Each *S* was seated in front of the lever manipulandum and knob manipulandum, and told:

> Please sit and remain in this chair until you are told the experiment is over. Before you, you see a lever and a knob.

When I say to you "go ahead," you may if you wish manipulate the lever; you may if you wish manipulate the knob; you do not have to manipulate the lever; you do not have to manipulate the knob; but you must not manipulate both the lever and the knob at the same moment. I repeat—you may if you wish manipulate the lever; you may if you wish manipulate the knob; you do not have to manipulate the lever; you do not have to manipulate the knob; but you must not manipulate both the lever and the knob at the same moment. Please do not attempt to take apart the apparatus. How you spend your time is up to you so long as you remain seated.

Do you have any questions about these directions?

All right. Go ahead.

The *E* then left the view of *S*, who was given the 5-min. operant period. After 5 min., *S*'s first lever manipulation activated the contingency, whereupon each 10 knob manipulations unlocked the lever until it was manipulated once. The contingency period lasted 5 min., and was followed by the 5-min. extinction period (unlocked knob and lever).

Results

As in Exp. II, the crucial *S*s were those undergoing response suppression and also having the higher operant probability of the instrumental response (knob manipulating) than contingent response (lever manipulating). As a partial control for a decline in the free-performance level of contingent responding, *S*s whose number of lever manipulations was zero for the last half of the operant period were excluded from the crucial group (it is suggested additional statistical controls, such as those employed in Exp. I, be used for future research). For the 12 *S*s failing to press the lever even once in the last half of the operant period, none of the three alternative hypotheses (see Exp. II) predicted the reinforcement effect. Eleven of these 12 *S*s decreased instrumental responding from the operant to contingency period, and each of 11 *S*s receiving the extinction period (the extinction period was not initiated until after the first few *S*s had been run) performed fewer knob manipulations than in the operant period. The median numbers of knob manipulations for the operant, contingency, and extinction period, respectively, were 3.2, 0.1, and 0.0. One additional *S* was excluded from the crucial group for having a higher operant duration (higher average probability) of lever manipulating than knob manipulating. The remaining 12 *S*s had a higher operant duration of knob manipulating than level manipulating (see Table 1) and also underwent response suppression, having performed more than one-tenth as many operant lever manipulations as knob manipulations. The response suppression hypothesis predicted the reinforcement effect, while the two other alternative hypotheses did not. A nonparametric statistical test sensitive to magnitude of change, the Wilcoxon signed-rank test, was utilized. In accordance with the response suppression hypothesis, there was a statistically significant increase in the number of knob manipulations (instrumental response) from the operant to contingency period, $T = 5$, $p < .005$, one-tailed test. Ten of the 12 *S*s increased their number of knob manipulations, the median number rising from 40.5 to 160. Similarly, there occurred a statistically significant increase in the number of knob manipulations from the operant to extinction period, $T = 10$, $p < .05$, one-tailed test. Of the 10 *S*s receiving the extinction period, 7 increased their number of knob manipulations above operant level, the median number of knob manipulations during the extinction period being 132 as again compared with 40.5 in the operant period. Table 1 lists the number of knob

TABLE 1

OPERANT DURATION OF KNOB MANIPULATING (INSTRUMENTAL RESPONSE) AND LEVER MANIPULATING (CONTINGENT RESPONSE), AND NUMBER OF KNOB MANIPULATIONS IN OPERANT, CONTINGENCY, AND EXTINCTION PERIODS FOR EACH CRUCIAL *S*

Item	Subject											
	1	2	3	4	5	6	7	8	9	10	11	12
Operant Duration of Responding (Sec.)												
Knob	92	30	46	56	49	63	23	132	109	104	68	72
Lever	85	9	11	24	31	37	6	60	49	32	27	25
Number of Knob Manipulations per Period												
Operant	62	17	18	25	60	23	22	34	47	162	131	153
Contingency	138	172	89	292	221	31	9	224	130	363	411	101
Extinction			142	165	123	3	19	245	116	257	156	34

Note.—Extinction period was instituted starting with the third crucial *S*.

manipulations for each *S* in each of the three periods.

DISCUSSION

Under the following conditions, an increase in instrumental responding is required if the individual is to have the opportunity of performing the contingent response as much in the contingency period as he would in the free-performance situation:

$$\frac{I}{C} \times \frac{O_c}{O_i} > 1$$

where I is the required amount of instrumental responding, C is the maximum amount of contingent responding subsequently allowed, O_c is the amount of free-performance contingent responding which would occur were there no instrumental requirement, and O_i is the amount of free-performance instrumental responding which would occur were there no instrumental requirement. The free-performance amounts of O_c and O_i are assessed in the operant period. The response units used to determine whether a particular individual undergoes response suppression are specified by the instrumental requirement; the unit for C being used also for O_c, the unit for I being used also for O_i. For example, for each *S* in Exp. I (where 1 bar press unlocked the wheel for 10 sec. at a time), C would be 10 sec. and O_c would be the number of operant seconds spent turning the wheel, I would be 1 bar press and O_i would be the number of operant bar presses. In Exp. IV (where 10 knob manipulations unlocked the lever for 1 manipulation), C would be 1 lever manipulation and O_c would be the number of operant lever manipulations, I would be 10 knob manipulations and O_i would be the number of operant knob manipulations.

Using the above formula, the specific conditions associated with the reinforcement effect in Premack's confirmatory probability-differential experiments (1959, 1963) may be derived as a special case of response suppression. In those particular experiments, each instrumental response allowed one contingent response; the ratio $I{:}C$ was therefore 1:1. Also, the individual instrumental response took about as long to perform as the individual contingent response (Premack, 1963, p. 88; 1965, p. 141); thus the ratio $O_c{:}O_i$ is changed little by expressing O_i and O_c as durations of responding or as numbers of responses. Placing Premack's instrumental requirement into the response suppression equation, it is seen that for each individual, response suppression occurred when the performance of the more probable response was made contingent upon the less probable response, but not when the performance of the less probable response was made contingent upon the more probable response.

The present set of experiments suggests the necessary and sufficient condition for reinforcement in the contingency situation is the animal's necessity to increase instrumental responding if it is to maintain contingent responding at the free-performance level.

REFERENCES

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.,* 1959, **66,** 219–233.

PREMACK, D. Predicting instrumental performance from the independent rate of the contingent response. *J. exp. Psychol.,* 1961, **61,** 163–171.

PREMACK, D. Reversibility of the reinforcement relation. *Science,* 1962, **136,** 255–257.

PREMACK, D. Rate differential reinforcement in monkey manipulation. *J. exp. Anal. Behav.,* 1963, **6,** 81–89.

PREMACK, D. Reinforcement theory. In D. Levine (Ed.), *Nebraska symposium on motivation, 1965.* Lincoln: Univer. Nebraska Press, 1965. Pp. 123–180.

PREMACK, D., & PREMACK, A. J. Increased eating in rats deprived of running. *J. exp. Anal. Behav.,* 1963, **6,** 209–212.

SCHAEFFER, R. W. The reinforcement relation as a function of the instrumental response base rate. *J. exp. Psychol.,* 1965, **69,** 419–425.

(Received July 7, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 351–355

RETROACTIVE INHIBITION AS A FUNCTION OF RESPONSE-CLASS SIMILARITY

MARK J. FRIEDMAN AND JAMES H. REYNOLDS
Colgate University

Postman, Keppel, and Stark (1965) have hypothesized that degree of retroactive inhibition (RI) occurring in an A-B, A-C paired-associates task is directly related to degree of response-class similarity in the A-B and A-C lists. The present experiment tested this hypothesis, employing an A-B, A-C paradigm in which the List 1 responses were CVCs and the form class of List 2 responses was either identical to (CVCs), similar to (CVs or CCCs), or different from (3-digit numbers) the class of responses employed in the A-B list. Results from a modified free-recall test of List 1 associations administered after A-C learning confirmed the hypothesis, showing that the greatest amount of RI occurred in the condition in which form class was identical in the 2 lists, with a gradual reduction in RI as the response classes became more dissimilar. The implications of these findings for existing theoretical explanations of RI were discussed.

Most investigations of retroactive inhibition (RI) in paired-associate (PA) learning have employed an A-B, A-C paradigm in which the responses in the two lists are of the same form class, e.g., adjectives, CVCs, etc. In a recent study, however, Postman, Keppel, and Stark (1965) varied the response *classes* in the A-B, A-C lists (adjectives for the B responses and letters for the C responses, or vice versa), and found less RI under these conditions than when the response class in the two lists was the same. The results were interpreted in terms of the unlearning hypothesis (Postman, 1961), which attributes RI to the unlearning of A-B associations assumed to occur when covert or overt responses from the first list are evoked during second-list learning and consequently extinguished because they are not reinforced. It was assumed that this unlearning of A-B associations, with resulting RI, occurred when the response classes in the two lists were the same; but that when response classes differed *S* did not evoke many first-list responses during second-list learning because such responding was irrelevant to a task so different in form class, and thus extinction and consequent RI did not occur. The latter part of this explanation utilizes Underwood and Schulz's (1960, p. 143) concept of a response selector mechanism, which assumes that when a task is being learned *S* restricts his range of responses to the category of response items relevant to that task.

On the basis of their findings Postman et al. have hypothesized that RI varies directly with the similarity of the response classes in the two lists of an RI paradigm. This principle extends Osgood's (1949) hypothesis of an inverse relationship between RI and response similarity when both the stimuli and the class of responses in the two lists are the same, and is a potentially important new addition to the full specification of the conditions under which RI occurs. To date, however, empirical evidence is limited to the single study comparing just two levels of response-class similarity. The present experiment, using an A-B, A-C paradigm followed by a modified free-recall (MFR) test to evaluate availability of first-list associations, compared the effects of four levels of response-class similarity—response classes in the two lists either identical (CVC–CVC), similar at two levels (CVC–CV or CCC), or different (CVC–3-digit numbers)—upon degree of RI.

METHOD

Subjects.—The *S*s were 50 undergraduate students taking introductory psychology at Colgate University, some of whom had had previous experience with paired-associate learning, but never with an RI task. Assignment to conditions was made as *S* appeared for the experiment according to a predetermined order in which each of the five conditions was assigned once at random before any condition was assigned again.

Design and materials.—The design consisted of four RI conditions, which differed from each other in degree of similarity of form class between first-list and second-list responses, and a control condition. The RI groups learned two PA lists (A-B, A-C) and after the second list were given an MFR test on the first set of associations. The control group learned just the A-B list, rested during the second-list learning period of the work conditions, and then received the MFR test.

The eight stimuli used in all lists were small color-disks (red, orange, yellow, green, blue, purple, brown, and black), approximately 7 mm. in diameter, punched from colored construction paper and pasted on the drum tape. For all groups, the first-list responses were CVCs selected from the 50–65% range on the Krueger (1934) association scale, with intralist similarity minimized as much as possible.

The second-list responses for one of the RI groups (Group CVC) were also CVCs, taken from the 35–50% range of Krueger's list, and representing a List 2 response class identical to that of the List 1 responses. Two other RI groups (Group CV and Group CCC) received as second-list responses either CVs of lowest rated frequency from the Underwood and Schulz (1960, Appendix F) tables, or CCCs rated at 80–90% association value on the Witmer (1935) scale. The second-list responses for the fourth condition (Group Num) were eight 3-digit numbers. The particular List 2 responses used for each RI condition were selected following pilot work which tested the learning equivalence of the varying response types at different association values, and indicated that the ones described above were roughly equivalent in terms of learning difficulty.

Three different orderings of the stimulus-response pairings were used in all lists to minimize serial learning of the responses and to provide three different starting positions.

Procedure.—The same procedure described by Postman et al. (1965) was used. Both the A-B and A-C lists were learned by the anticipation method at a 2:2 sec. rate, with a 4-sec. intertrial interval. The A-B list was presented to a criterion of one perfect recitation. All *S*s were then given the instructions for the second task during approximately a 1-min. period following the attainment of criterion. For each RI group, the second list was then presented for 20 anticipation trials. The control group was given the MMPI as a filler task for 12 min., equivalent to the time taken by the RI groups to complete the 20 trials of the List 2 task. At the end of the interpolated activity all groups were given the MFR test, which was unpaced. For the control group, *S* was shown the stimuli one at a time with instructions to recall and spell aloud the correct response. For the RI groups a variation of the MFR test was used, which involved the presentation of each of the eight stimulus terms along with the correct List 2 response, and *S* was instructed to recall and spell aloud the appropriate response term from the first list. The order of the stimulus terms on the MFR test corresponded to the order of the pairs on the trial following the last List 2 trial administered.

RESULTS

First-list learning.—Means and *SD*s of number of trials to criterion on initial A-B learning for all conditions are presented in the first two rows of Table 1. A simple analysis of variance showed that differences among the means were not significant, $F\ (4, 45) = .183,\ p > .05$.

Second-list learning.—Rows three and four of Table 1 present the means and *SD*s for number of trials to mastery of the A-C task for the four RI conditions. The differences among the means on this measure were not sig-

TABLE 1

Learning and Recall Measures for the Five Conditions

	List 2 Response-Class Groups				
	CVC	CCC	CV	Numbers	Control
List 1					
Trials to Criterion					
$\bar{X}$	22.8	21.5	20.8	19.0	22.7
SD	9.6	12.3	13.6	10.4	8.1
List 2					
Trials to Criterion					
$\bar{X}$	12.5	10.6	9.9	13.8	—
SD	4.4	5.1	4.6	4.0	—
MFR					
Number Correct					
$\bar{X}$	3.4	5.0	5.1	6.7	7.7
SD	2.1	2.2	1.7	1.0	.46

nificant, $F\ (3,\ 36) = 1.36,\ p > .05$. The learning curves for these groups over the 20 List 2 trials also indicated that the four groups were nearly equivalent in List 2 performance, although the performance of Group Num was slightly below the other groups in the middle trials. These data suggest that the use of different association levels of the four response classes, as determined from the pilot study, was successful in providing second-list tasks of equivalent learning difficulty.

A count of the interlist intrusions made by each group during List 2 learning revealed that three *S*s in Group CVC made a total of eight overt interlist errors, while no interlist intrusions were made in Groups CV, CCC, or Num. Although these numbers are small, they do suggest the presence of some interlist response competition when the form class of the two lists was identical, but do not show evidence of competition at each of the various levels of response dissimilarity.

First-list recall.—The last two rows of Table 1 present means and *SD*s of correct List 1 responses on the MFR test. Inspection shows that each of the four RI conditions recalled fewer List 1 responses than the control group, and that the mean number of correct responses declines as the similarity between the List 1 and List 2 response classes increases. An analysis of variance performed on these means was highly significant, $F\ (4,\ 45) = 9.32,\ p < .01$.

Further evaluation of differences in RI among the groups was made by a series of orthogonal individual comparisons (Winer, 1962, pp. 65–77). The comparison made between Group Num and the control group was not significant, $F\ (1,\ 45) = 1.68,\ p > .10$, indicating that the second-list task in which responses differed in form class from List 1 responses did not yield significant RI relative to a group receiving no List 2 learning. However, the difference between these two groups combined and Group CV was significant, $F\ (1,\ 45) = 9.90,\ p < .01$, indicating the presence of RI for a group receiving similar form classes of response on the first and second lists. Since the means for Group CV and Group CCC were nearly identical, no statistical comparison between these groups was made. The final comparison was between Group CCC and

Group CVC. The difference between these groups was significant, F (1, 45) = 4.31, $p < .05$, demonstrating that RI for the condition in which List 1 and List 2 response classes were identical was reliably greater than for a condition in which response classes were similar.

The total frequencies of intralist errors were 3, 3, 7, 3, and 2 for Groups CVC, CCC, CV, Num, and the control group, respectively. No interlist errors occurred at recall in any of the RI groups.

DISCUSSION

The present data show large and significant RI in an A-B, A-C paradigm when the B and C responses are from the identical form class, less but still significant degrees of RI when the responses are from similar form classes, and an absence of significant RI relative to a rest control when the C responses differ in category of form class from the responses in the first list. These findings confirm the previous results reported by Postman et al. (1965), and support their hypothesis that "the more dissimilar the classes of responses attached to the same stimuli the more likely it becomes that the successive response systems will maintain their strengths independently of each other [p. 118]." Thus similarity is reaffirmed as a major variable contributing to RI; but in contrast to the inverse relationship existing between RI and the similarity of responses within the same class (Osgood, 1949), there appears to be a direct RI-similarity relationship for the dimension of response-class similarity.

A theoretical explanation of these results may be obtained by employing the same unlearning hypothesis and selector mechanism concept used by Postman et al. in explaining their findings, but with some modification in the assumed properties of the latter concept. Interpreted within the framework of the unlearning hypothesis, the present findings imply that overt or covert List 1 response-evocation tendencies (and therefore extinction of first-list responses) during second-list learning decreased in a regular fashion as the similarity of the responses in the A-B and A-C lists decreased. While the interlist intrusion data for List 2 learning fails to offer evidence of a gradual decrease in Overt List 1 intrusions as response-class similarity decreases, the total number of intrusions made is too small to justify interpreting this failure as negative evidence for the hypothesis. Instead, the low number of overt intrusions may be taken as indicating that the hypothesized relationship between response-evocation tendencies and response-class similarity applies primarily to covert List 1 responses which are made at the time of List 2 learning but not vocalized.

As Postman et al. have shown, the concept of a selector mechanism (Underwood & Schulz, 1960) appears to be essential to an unlearning explanation of response-class effects upon RI, since it offers a theoretical reason for the failure to extinguish A-B associations during A-C learning when response classes differ. This mechanism is interpreted as predicting that when response classes in the two lists of an A-B, A-C paradigm differ S will not make interlist intrusions during second-list learning because the potentially intruding (B) responses are in a different category from those of the second list, and thus A-B associations will not undergo extinction. To account for the present data, however, it is necessary to assume that the mechanism does not select out equally in all situations the category of responses being learned, but instead operates as a *tendency* to select the appropriate category of responses, this tendency varying on a continuum with the similarity between the category of items in the list being learned and the category of the potentially intruding responses. Thus the direct relationship obtained between RI and response-class similarity can be explained adequately by the unlearning hypothesis, but such an explanation would seem to require a selector-mechanism concept which functions on a continuous, rather than an either-or, basis.

REFERENCES

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152–179.

POSTMAN, L., KEPPEL, G., & STARK, K. Unlearning as a function of the relationship between successive response classes. *J. exp. Psychol.*, 1965, **69**, 111–118.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received July 8, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 356–362

ON THE LENGTH-DIFFICULTY RELATION IN SKILL PERFORMANCE[1]

JAY SWINK, DON TRUMBO, AND MERRILL NOBLE

Kansas State University

Sequence length in an irregular step function tracking task was varied in order to evaluate the effect of sequence length on acquisition and retention of a motor skill. A 5 × 2 × 2 × 2 factorial design was used with 5 sequence lengths, 2 degrees of task coherence, 2 training criteria (equal practice vs. equal repetitions), and 2 long-term retention intervals. The results support Deese's contention, based on verbal serial learning, that repetitions per item do not increase with length of the series, although total learning time and total number of repetitions do increase approximately as suggested by Thurstone's exponential equation.

The classical learning literature has maintained that the longer a task is the more difficult it is to learn. This is commonly referred to as the length-difficulty relationship and has been an area of investigation for verbal learning researchers since Ebbinghaus. The nature of this relationship has generally been found to be that lengthening lists disproportionately increases difficulty of learning, and hence, time to learn (McGeoch & Irion, 1952). However, Deese (1958), in reviewing Lyon's (1917) monograph, found evidence indicating that while longer lists increased total time to learn, the time per item did not increase appreciably and, moreover, that fewer repetitions per item were required to learn a list of 200 items than a list of 16 items, even though the total number of repetitions was greater for the longer list. Deese concluded that, "the results of the available experimental evidence clearly suggest that the amount of material does not necessarily disproportionately increase difficulty of learning as has commonly been supposed both by learning theorists and those only casually interested in learning [Deese, 1958, p. 212]." Deese's analyses agree in general with the earlier work of Thurstone (1930), who stated (*a*) that the time required to learn varies approximately as the 3/2 power of the length of the list to be learned and results in a positively accelerated function with increasing list lengths, and (*b*) that the number of repetitions required for learning increases as the square root of the number of items in the list, with a correction for the attention span, resulting in a function which is negatively accelerated.

There is only scanty evidence as to the effects of task length upon motor learning. Scott and Henninger (1933) found similarity of the length-difficulty relationships in verbal and motor serial tasks (finger mazes). More recently, Trumbo, Noble, and Ulrich (1965) examined the effects of different-sequence lengths in an irregular step function, sequential tracking task, and found no difference between sequences of 5 and 10 units (targets) in length but significantly poorer performance with sequences of 15 targets. However, these three groups received an unequal number of repetitions of the sequence since they received an equal amount of practice time. When subsequent comparisons of performance were made after each group had received 400 repetitions, the error scores were nearly equal.

---

[1] This report is based on a Master's thesis submitted to the Graduate School, Kansas State University, by the first author. The research was supported by the National Aeronautics and Space Administration under Grant No. NsG 606.

In spite of the lack of evidence on the length-difficulty relationship, it usually is thought to be an important factor. Adams (1964), for example, stated that, "the number and type of S-R sequences, and their rate, that *S* can process per unit of time is fundamental to a theory of a quantified sort in this area [p. 197]." It is not clear, however, whether the course of either acquisition or retention is altered by length of a motor task, or, if such alterations exist, whether they would be equivalent to those found in verbal serial learning.

In the present study the length of a sequence of targets in pursuit tracking was varied in a manner similar to that used when varying list length in a verbal serial learning task. A second independent variable, task predictability or coherence, was also investigated in an attempt to assess the possible interaction of sequence length and task organization as measured either by criterion performance or by spatial-temporal response patterning. Previous findings involving this task organization variable (Trumbo, Noble, Cross, & Ulrich, 1965) have indicated that both acquisition and retention performance was positively related to the proportion of repeating elements in a sequence.

METHOD

Subjects

The *S*s were 120 right-handed undergraduate male students between 17 and 24 yr. of age. They received credit in introductory psychology courses or were paid for the services.

Apparatus

The apparatus was essentially a replica of that described elsewhere in detail (Trumbo, Eslinger, Noble, & Cross, 1963). A programming and scoring unit controlled the input to two essentially identical *S* booths. Target programs were punched on paper tape, read out by a commercial tape reader, converted by a digital-to-analog converter and a flip-flop circuit to analog voltage and displayed on a CRT as a ½-in. vertical hairline. By moving a lateral arm control, pivoted at the elbow and connected to a potentiometer, *S* controlled a similar vertical line. This second line, the cursor, was displayed below the target on the CRT with an overlap of ⅛ in. The control-to-display ratio was 11.25° arc to 2 cm. of cursor displacement.

Scoring was accomplished by an operational amplifier manifold which obtained the absolute difference between the target and cursor and integrated that difference over the trial duration. Input, output, and momentary error were recorded simultaneously for selected trials on a magnetic tape, and integrated error was recorded by *E* on each trial from a voltmeter.

Experimental Conditions

The experimental design was a $5 \times 2 \times 2 \times 2$ complete factorial with five sequence lengths, two degrees of predictability, two training criteria and two retention intervals. Six *S*s were randomly assigned to each of the 20 training conditions. Each group was subdivided into two subgroups, matched on the basis of integrated error scores, for retention, resulting in 3 *S*s in each of the 40 retention conditions.

Target sequence.—The target sequence for all conditions was an irregular step-function input. The target line moved once per sec. from 1 of 15 equidistant positions to another within the middle 4 in. of the horizontal axis of a 5 in. CRT. A series of 48 step changes constituted a trial and trials were separated by a 12-sec. rest period. A 2-sec. warning buzzer signaled the beginning of each trial.

Sequence length.—Sequence length refers to the number of targets in a basic sequence repeated both within and throughout all trials. Sequence lengths of 8, 12, 16, 24, and 48 targets were employed. Since trial duration was constant for all conditions, a sequence of 8 targets was repeated 6 times per trial, a sequence of 12 repeated 4 times per trial, etc. The first target of each sequence appeared in the middle of the scope and this position was used exclusively to cue the beginning of a sequence.

Task predictability.—Two conditions of task predictability were employed. In one condition the targets appeared in the same order and position with each repetition of the sequence. In the other, every fourth target position was selected at random each time the sequence was repeated. Thus, in one condition, 100% of the targets were repeated in each repetition and in the other, 75% were repeated.

For sequence lengths of 8 and 12, the basic (100%) sequence was selected randomly without replacement from the 14 possible target positions. For the remaining sequence lengths, replacement occurred in the random selection only after all 14 possible positions had been used, and then no position could be repeated immediately. For all sequences, the initial target position was in the center of the CRT. Three different basic target sequences were constructed for each of the 5 sequence lengths, and an equal number of *S*s assigned to each of these.

Training criteria.—Under the criterion of equal practice time each *S* received 20 trials per day for 5 consecutive days. Under the criterion of equal repetitions each *S* received 360 repetitions of a given sequence, except for sequences of 48 targets for which only 180 repetitions were given.

Retention interval. — Performance was measured after 3 mo. or 5 mo. without practice. All *S*s received a total of 40 retention trials, with a 5-min. rest after the first 20 trials.

Procedure.—Two *S*s were scheduled at a time and tested under the same experimental conditions. Detailed instructions were given to all *S*s, including a description of the task, the method of performance scoring, and the identification of the principal sources of errors via sample oscillographic records. The efficiency of anticipation and rapid primary movements was explained and it was emphasized that the center position of the scope was the cue for the beginning of the sequence.

Knowledge of results was provided via an intercom to *S*s after alternate trials by stating the total integrated error accumulated on the preceding trial.

Performance measures.—Performance measures included absolute error integrated over the duration of each trial, and a temporal index consisting of the algebraic sum of all the leads and lags (scored to the nearest 50 msec. as the interval between the target displacement and the initiation of the primary-pursuit movement) in a trial, divided by the total number of targets in the trial (48). The latter scores were obtained by hand scoring from oscillographic records. The oscillographic recorder employed mirror galvanometers; hence target and cursor signals could be superimposed to facilitate scoring.

RESULTS

Integrated error.—The data for the conditions having an equal amount of practice time are presented in Fig. 1. A repeated-measure analysis of variance indicated that Predictability (P) and Sequence Length (SL) were both highly significant ($p < .01$) but the P × SL interaction was not significant. The practice effect analyzed as Blocks (B) was also highly significant ($p < .01$) as were the interactions of Blocks with the other two independent variables (B × P, B × SL, B × P × SL) indicating differential acquisition rates across trials. A mean separation test (Fisher's lsd) performed on the last block of 10 acquisition trials indicated that for the predictable patterns only the 48-target sequence differed significantly from other conditions. For the variable patterns all sequence lengths differed significantly ($p < .01$) from each other except the immediately adjacent pairs.

The data for conditions equated for number of repetitions of a sequence are summarized in Fig. 2. An analysis of variance indicated again that Predictability was highly significant ($p < .01$) but neither Sequence length nor P × SL were significant. The practice effects (Blocks) and all interactions involving this repeated measure were again highly significant ($p < .01$), indicating a differential change over trials.

Temporal index.—Mean lead-lag scores are presented in Fig. 3 for 3

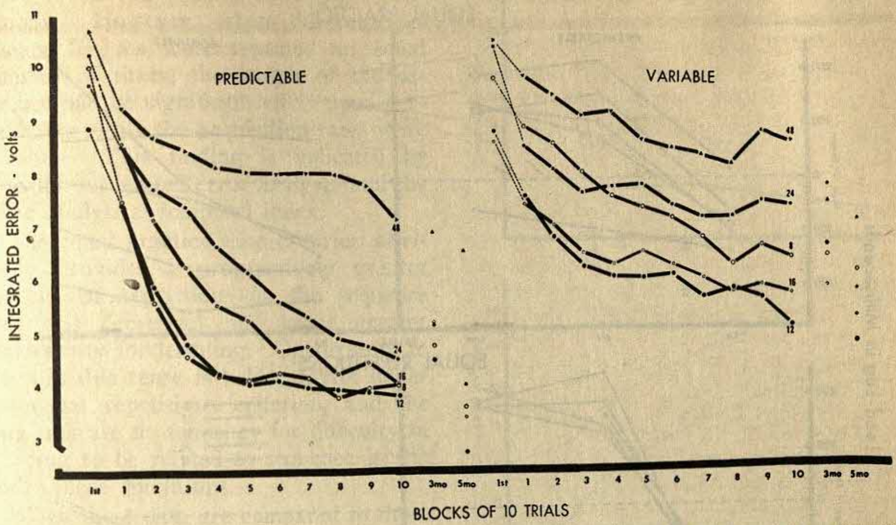


FIG. 1. Integrated error as a function of sequence length with both Predictable (100%) and Variable (75%) sequences. (All *S*s were run to an equal practice time criterion—100 48-sec. trials.)

representative* *S*s from each group. There was a marked tendency for *S*s having completely predictable sequences to shift from lagging behind the target to leading the target as practice continued. Inspection of Fig. 3, however, indicates little difference in this tendency to change from lagging to leading the target between sequence lengths when all groups were given an equal number of repetitions of the sequence, but a considerable spread among the

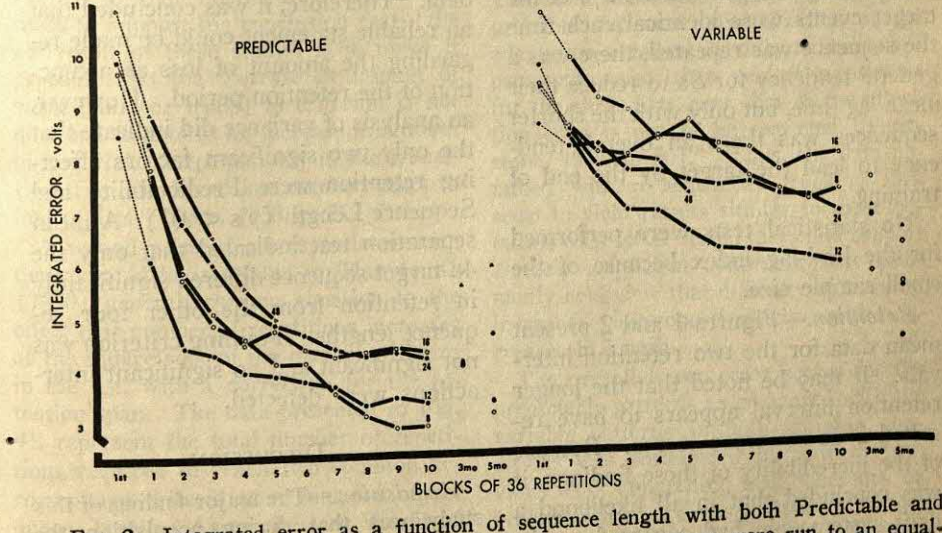


FIG. 2. Integrated error as a function of sequence length with both Predictable and Variable sequences. (All *S*s except those in the 48-target sequence were run to an equal-repetitions criterion—360 repetitions.)

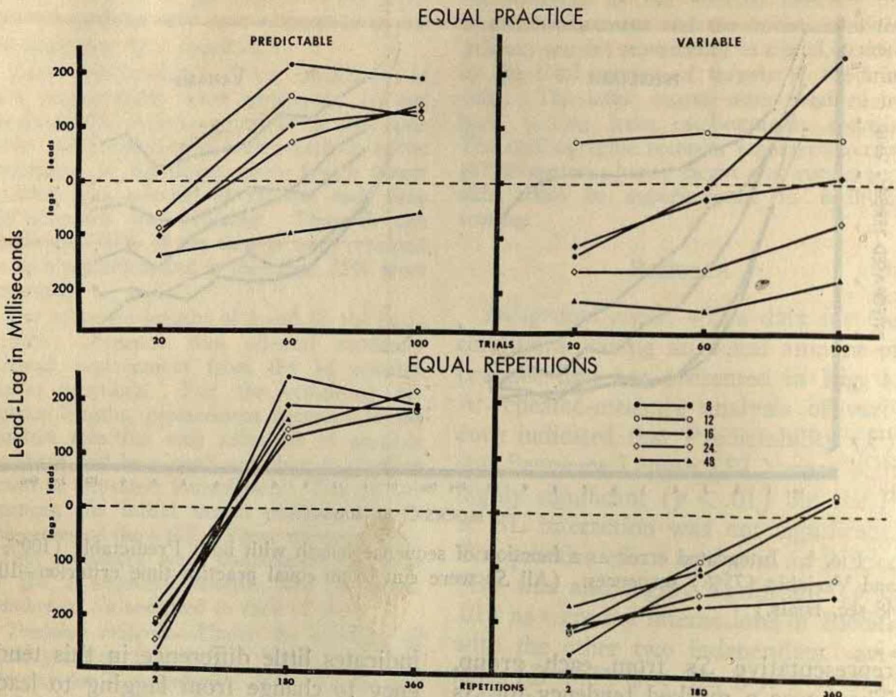


FIG. 3. Lead-lag data for all conditions at three levels of training. (Data points represent means for three *S*s representative of each condition.)

several groups when the amount of practice time was held constant.

With sequences in which 75% of the target events were identical each time the sequence was repeated, there was a general tendency for *S*s to reduce their mean lag time, but only with the shorter sequences was there an overall tendency to lead the target by the end of training.

No statistical tests were performed for the lead-lag index because of the small sample size.

Retention.—Figures 1 and 2 present mean data for the two retention intervals. It may be noted that the longer retention interval appears to have resulted in lower error scores. Because of the incredibility of these findings, it was concluded that in all likelihood a drift system gain had occurred resulting in a constant lowering of error scores at 5 mo., although no evidence for this could be found in calibration data. Therefore, it was concluded that no reliable statement could be made regarding the amount of loss as a function of the retention period. However, an analysis of variance did indicate that the only two significant factors affecting retention were Predictability and Sequence Length (p's $< .01$). A mean separation test indicated that only the 48-target sequence differed significantly in retention from the other four sequence lengths. Training criterion was not significant and no significant interactions were detected.

DISCUSSION

Acquisition.—The major finding of this study was that during acquisition the length of the sequence had a significant effect upon performance and acquisition rates under the equal practice time conditions. However, when different sequence lengths were repeated an equal number of times, the length of the sequence had no significant effect upon performance; only the acquisition rates were affected. This finding is indicated by both the integrated error analyses and the more analytical temporal index.

The equal practice time criterion obviously provides a progressively greater number of repetitions as the sequence length is decreased, and hence greater opportunity for learning. Opportunity to learn in this sense is held constant under the equal repetitions criterion, and the data indicate no tendency for difficulty of learning to be related to sequence length under these conditions.

When these data are compared to those from verbal learning the results appear to be similar. It will be recalled that when Deese (1958) replotted Lyon's (1917) data, the total number of repetitions *increased* with the longer lists, but the number of repetitions per item required to reach a common criterion *decreased* as the length of the list was increased. In Fig. 4A, repetitions per target required to reach two different integrated error criteria are plotted for the sequences of varying lengths. The function is similar to that obtained by Deese and tends to support his conclusion that, "the net result of these studies is to make it probable that the increase in amount of work [time or trials] to criterion is not disproportionate to an increase in amount of material, except possibly in the transition from the immediate memory span to longer lists [Deese, 1958, p. 212]."

A second point of correspondence with the verbal data is related to Thurstone's (1930) general learning equation to the effect that number of repetitions increases as the square root of the number of items in the list, with a correction for the attention span. The data presented in Fig. 4B represent the total number of repetitions required to reach two common error-voltage criteria as a function of sequence length. As can be seen, the fit of the empirical data to Thurstone's theoretical equation is quite good with the exception of the eight-target sequence. Thurstone indicated that, "it is quite probable that our rationalization of the adaptation constant is only a rough approximation so that perfect agreement for the short lists near that of the attention span is not to be expected [Thurstone, 1930, p. 50]." Thus, the data obtained with a sequential tracking task seem to yield results similar to those obtained with serial verbal tasks and support Deese's skepticism concerning the commonly held view that difficulty of learning increases disproportionately with increases in length.

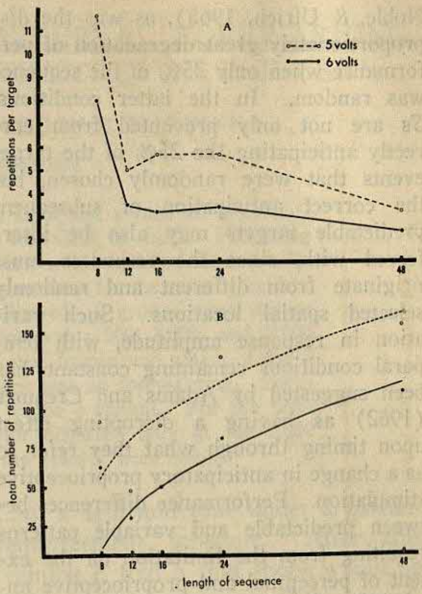


FIG. 4. Repetitions per target (A) and total repetitions of the sequence (B) required to attain two integrated error criteria as a function of sequence length. (Data points are obtained data, while the smoothed curves in Fig. 4B represent Thurstone's, 1930, theoretical equation.)

The overall lower error scores for the predictable patterns as compared with the variable patterns in both training conditions was anticipated in view of previous work concerning task coherency or predictability (Cross, 1966; Noble, Trumbo, Ulrich, & Cross, 1966; Trumbo, Noble, & Ulrich, 1965), as was the disproportionately great degradation of performance when only 25% of the sequence was random. In the latter conditions, *S*s are not only prevented from correctly anticipating the 25% of the target events that were randomly chosen, but the correct anticipation of subsequent predictable targets may also be interferred with, since the responses must originate from different and randomly selected spatial locations. Such variation in response amplitude, with temporal conditions remaining constant, has been suggested by Adams and Creamer (1962) as having a disrupting effect upon timing through what they refer to as a change in anticipatory proprioceptive stimulation. Performance differences between predictable and variable patterns resulting from the limitation on the extent of perceptual and proprioceptive anticipation was very evident in the temporal index data (Fig. 3).

Retention.—The principal finding was that training criterion was not a significant variable during retention.

Sequence length, which was a significant variable only when practice time was equated, was also a significant overall variable at retention. However, this was due primarily to the 48-target sequences which did not receive equivalent training even under the equal repetitions criterion (only 180 repetitions).

The variable of predictability, which was significant in all conditions of acquisition, was also significant at retention. Apparently the difference between 100% and 75% predictable targets is enough to cause significant performance differences even after 3 or 5 mo. of no practice.

REFERENCES

ADAMS, J. A. Motor skills. *Ann. Rev. Psychol.,* 1964, **15,** 181–202.

ADAMS, J. A., & CREAMER, L. R. Proprioceptive variables as determiners of anticipatory timing behavior. *Hum. factors,* 1962, **4,** 217–222.

CROSS, K. Discrete tracking proficiency as a function of temporal, directional, and spatial predictability. Unpublished doctoral dissertation, Kansas State University, 1966.

DEESE, J. *The psychology of learning.* New York: McGraw-Hill, 1958.

LYON, D. O. *Memory and the learning process.* Baltimore: Warwick & York, 1917.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, 1952.

NOBLE, M., TRUMBO, D., ULRICH, L., & CROSS, K. Task predictability and the development of tracking skill under extended practice. *J. exp. Psychol.,* 1966, **72,** 85–94.

SCOTT, T. C., & HENNINGER, L. L. The relation between length and difficulty in motor learning; a comparison with verbal learning. *J. exp. Psychol.,* 1933, **16,** 657–678.

THURSTONE, L. L. The relation between learning time and length of task. *Psychol. Rev.,* 1930, **37,** 44–53.

TRUMBO, D., ESLINGER, R., NOBLE, M., & CROSS, K. A versatile electronic tracking apparatus (VETA). *Percept. mot. Skills,* 1963, **16,** 649–656.

TRUMBO, D., NOBLE, M., CROSS, K., & ULRICH, L. Task predictability in the organization, acquisition, and retention of tracking skill. *J. exp. Psychol.,* 1965, **70,** 252–263.

TRUMBO, D., NOBLE, M., & ULRICH, L. Number of alternatives and sequence length in acquisition of a step-function tracking task. *Percept. mot. Skills,* 1965, **21,** 563–569.

(Received July 8, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 363–369

EFFECTS OF INTERSTIMULUS INTERVAL, FOREWARNING, AND INSTRUCTIONS ON EXTINCTION OF THE GSR[1]

DELOS D. WICKENS AND GORDON B. HARDING[2]

Ohio State University

Ss were conditioned concurrently to a light and a tone on a partial-reinforcement schedule. The response was GSR to shock. Following conditioning they were told that no shock would be administered with 1 of the 2 stimuli, implying the usual procedure for the other stimulus. 4 extinction trials to each stimulus followed. 2 groups were run at each of the following interstimulus intervals: 500, 1,000, and 2,000 msec. For 1 group at each interval S was forewarned which of the 2 stimuli would occur; for the other there was no forewarning. Forewarned groups behaved differentially to the 2 stimuli. For the non-forewarned groups differential behavior occurred only at the 2,000 interval. Several interpretations of the results were made.

Although there is much evidence that responses following GSR conditioning will be decreased if *S* is told that no shock will be delivered thereafter, considerable inconsistencies in the details of this effect are to be found. Thus Silverman (1960) and Grings and Lockhart (1963) do not agree on the effects of instructions at a long interstimulus interval, while Wickens, Allen, and Hill (1963) and Bridger and Mandel (1965) disagree on the immediacy of the effect of the instructions. Such disparity suggests that the effect of instructions is not a simple one and that it interacts with other variables. In this research a within *S* design was used to evaluate the effect of instructions and forewarning of the nature of the succeeding CS. The results of this first experiment suggested that the effect of instructions might also be dependent on the CS-UCS interval. A second experiment was done to determine if this was the case.

EXPERIMENT I

The general procedure consisted of conditioning *S* concurrently to two CSs, a light and tone under a partial-reinforcement schedule. After this training *E* told *S* that shock would no longer accompany one of the stimuli while implying that it would continue to be associated with the other. Nonreinforced presentations of both CSs followed with half of the *S*s being forewarned of the nature of the CS next to occur.

Method

Subjects.—The *S*s were 32 students from the introductory psychology classes.

Apparatus.—The time relations and duration of the CS and UCS were controlled by Tektronix equipment and calibrated by a Hunter interval timer. A 1,000-cps tone (rise time 100 msec., 55 db. re .0002 dynes cm^2) served as one CS and was generated by a Jackson oscillator. The light CS came from an aperture, 3-in. diameter, located approximately 3 ft. in front of *S* at eye level and was psychologically equated to the tone in intensity. The shock UCS was produced by an Applegate constant current shock apparatus No. 125. The *S* was run in an IAC room, located inside the experimental room, and wore Willson sound barrier ear muffs containing earphones which delivered the tone. The booth was illuminated by a neon 40-w. bulb located above and to the right of *S*'s head. Two 7½-w. bulbs, one red, the other green, were located on the back of the booth, 1 and 2 ft., respectively, above *S*'s head. The illumination of one or the other

---

[1] This research was supported in part by United States Public Health Service Grant No. NH-08423.

[2] Now at New Mexico State University.

of these lights was used to indicate to *S* which CS was to occur next.

The GSR was measured by a Fels dermohmmeter using zinc electrodes and a jelly composed of bacto agar, zinc sulfate, and distilled water.

Procedure.—(*a*) Conditioning: After *S* entered the experimental room and was seated, the GSR electrodes were attached to the palmar and dorsal side of his right hand and the shock electrodes to the middle and index fingers of his left hand. The *S* was told that he would receive a series of lights and tones paired with a mild shock to his fingertips, but that there was nothing for him to do except remain alert. The *E* then closed the booth door and presented three shock adaptation trials with successive intensities of 1.0, 1.2, and 1.4 ma.

One-half of the *S*s received the tones and lights in the opposite order of the other half, the presence of shock on any given trial being the same for all *S*s. Each of the two stimuli was paired with the UCS eight times and presented alone five times. The intertrial interval varied between 15 and 70 sec. with a mean of about 40 sec. The shock stimulus was 100 msec. in duration and was increased from 1.0 ma. at the beginning of shock adaptation to 2.5 ma. in steps of .2 ma. and then .1 ma. The CS-UCS interval was 1,000 msec., the CS terminating with the shock.

(*b*) Extinction: Following the conditioning procedure, *E* gave all *S*s the following instructions:

> I am going to add a new condition to the experiment. Now, before each trial, a light will come on behind you and produce a red or a green glow on this front wall. Do not turn around to look directly at these lights.

Then, one-half of the *S*s, the Unwarned group, was told:

> This condition is to determine the effect of color on the efficiency of conditioning. In addition to this change, that is, the use of the colors, there will be another change in procedure. The tone (light) will no longer be paired with shock. The procedure for the light (tone) will remain the same as in the first part of the experiment. Now, can you repeat these instructions for me?

The Forewarned group was told:

> In addition to this change, that is, the use of the colors, there will be another change in procedure. The tone (light) will no longer be paired with shock. Green will always come before the tone (light) is presented. The procedure for the light (tone) will remain the same as in the first part of the experiment. Red will come on before the light (tone) occurs. Now, can you repeat these instructions for me?

The instructions were read again for *S*s not able to repeat the contingencies therein.

The *E* then left the booth and presented the Threat (shock implied) and No-Threat (no shock implied) stimuli in an ABBA-BAAB order or its converse with a mean intertrial interval of 38 sec. varying between 20 and 60 sec.

Each Unwarned *S* was exposed to only one colored light throughout extinction. For the Forewarned *S*s, the red light appeared prior to the Threat CS and the green light prior to the No-Threat CS. For all *S*s, the colored light appeared at least 15 sec. prior to the onset of the CS and terminated approximately 5 sec. thereafter.

The GSR was recorded visually from a dial on the face of the Fels dermohmmeter and was defined as the response which began within approximately 2 sec. of stimulus presentation. The scores were transformed according to the formula $[(R_b - R_a)\ 10^{-2}]^{\frac{1}{2}}$ where R_b is the resistance at the moment of initiation of the response and R_a is the resistance at the peak of the response. The visual recording was checked against an Esterline-Angus recorder.

Results

Response to the CS.—The data points in extinction were obtained by dividing each extinction CR by the mean CR to that same CS on its last three nonreinforced presentations during acquisition. Such a ratio score leaves the extinction data uncontaminated by differences in responsiveness to the two CSs during acquisition.

Figure 1 shows the mean performance of the Forewarned and Unwarned *S*s to the Threat and No-Threat CSs over Extinction tests. An Extinction test consists of an *S*'s CR to a Threat and No-Threat CS presented successively. An analysis of variance was conducted with Foreknowledge and

Modality of the CS as between-*S* variables and Instructions (Threat or No-Threat CS) and Extinction tests as within-*S* variables. The Newman-Keuls procedure was used for the a posteriori tests of the means.

The main effect of Instructions was significant, $F\ (1, 28) = 12.28,\ p < .01$, as was Extinction Tests, $F\ (3, 84) = 4.30,\ p < .01$, and their interaction with Foreknowledge: Instructions × Foreknowledge, $F\ (1, 28) = 8.38,\ p < .01$, Extinction Tests × Foreknowledge, $F\ (3, 84) = 3.81,\ p < .05$, and the triple order interaction, Instructions × Foreknowledge × Extinction Tests, $F\ (3, 84) = 3.80,\ p < .05$.

Analysis of the means of the Instructions × Foreknowledge × Extinction Tests interaction revealed that the Forewarned *S*s made a significantly larger CR to the Threat CS than to the No-Threat CS on every Extinction Test except the first. The Unwarned *S*s failed to respond differentially to the two CSs throughout extinction. Further, Forewarned *S*s made a greater CR to the Threat CS and a smaller CR to the No-Threat CS than did the Unwarned *S*s, significantly so on Extinction Tests 4 and 3, respectively. A significant decrement in response over Extinction Tests by the Forewarned *S*s to the No-Threat CS and by the Unwarned *S*s to the Threat CS also contributed to the triple-order interaction.

Regardless of the condition of Foreknowledge, differential performance to the two CSs was more likely, $F\ (1, 28) = 4.84,\ p < .05$, if the No-Threat CS was a tone than if it was a light.

Base level.—The mean base levels of the Forewarned *S*s at the time of the first and last presentation of the Threat and No-Threat CS were compared and found nonsignificant, $t\ (31) = .0055$. The mean base level prior to Threat CS onset for these trials was 48,975 ohms, while prior to No-Threat CS onset it was 49,825 ohms.

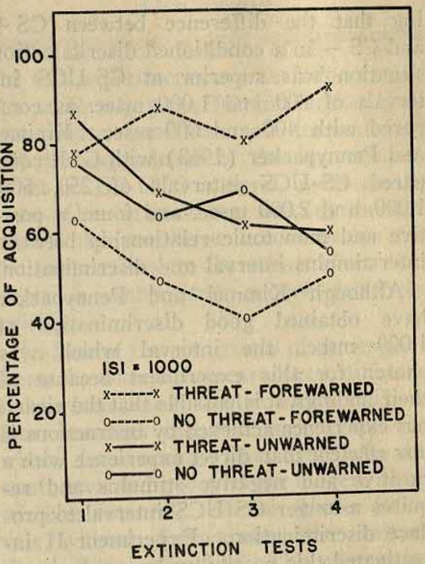


FIG. 1. Mean performance of the forewarned and unwarned *S*s to the threat and no-threat stimuli at the 1,000-msec. interval.

Discussion

The results of this experiment have clearly indicated that under conditions of forewarning *S*s respond differentially to stimuli designated as threatening and nonthreatening by previous instructions. Without such forewarning, no differentiation of responding occurred. Clearly, verbal instruction can be effective or ineffective depending upon the presence or absence of other variables, in this case forewarning, amount of extinction experience, and the modality of the stimuli mediating the instructions.

The significant Instructions × Foreknowledge interaction suggests that the failure to respond differentially in the forewarned group resulted from a failure in some aspect of an inhibitory process rather than a failure of *S* to accept the instructions. This interpretation becomes especially plausible in the light of data on differential conditioning and interstimulus interval. Hartman and Grant (1962) have demonstrated for eyelid conditioning that the difference between CS + and CS − in a conditioned discrimination situation was superior at CS-UCS intervals of 800 and 1,000 msec. as compared with 400 and 600 msec. Kimmel and Pennypacker (1963) with GSR compared CS-UCS intervals of 250, 500, 1,000, and 2,000 msec. and found a positive and monotonic relationship between interstimulus interval and discrimination.

Although Kimmel and Pennypacker have obtained good discrimination at 1,000 msec., the interval which was chosen for this experiment because of their findings, it is possible that the vicarious experience achieved by instructions is less efficient than direct experience with a positive and negative stimulus and requires a longer CS-UCS interval to produce discrimination. Experiment II investigated this possibility by employing a group with a CS-UCS interval of 2,000 msec. and a group with an interval of 500 msec. in order that a relatively broad range of CS-UCS intervals could be sampled.

EXPERIMENT II

Method

Subjects.—The 64 *S*s of Exp. II, 32 in each CS-UCS interval group, were drawn from the same population as *S*s of Exp. I. The *S*s in the two groups of this experiment were run concurrently during the last week in November and the first 2 wk. of December as compared with the last week in October and the first week in November for Exp. I.

Procedure.—The procedure in this experiment was identical with that of Exp. I save for CS-UCS interval which was 500 and 2,000 msec.[8]

Results

The data points during the Extinction tests were obtained using the same formula as in Exp. I. Separate analyses of variance were performed on the data from the Forewarned groups and the Unwarned groups as the variation of major interest was expected from the Unwarned groups.

Forewarned groups.—(*a*) Response to the CS: Figure 2 shows the mean CR magnitude to the Threat and No-Threat CS at the 500 and 2,000 msec. CS-UCS intervals. The main effects of Instructions, $F\ (1,\ 28) = 43.37,\ p < .01$, and Interstimulus Interval, $F\ (1,\ 28) = 6.14,\ p < .05$, were significant with no interaction. A Newman-Keuls test of the means of Fig. 2 shows that, while the mean CR to the Threat CS is significantly greater than the mean CR to the No-Threat CS at both the 500 and 2,000 interstimulus intervals, the mean CR magnitude to both the Threat and No-Threat CS decreased significantly as the interstimulus interval became long. The effect of Extinction tests was not significant, $F\ (3,\ 84) < 1$.

(*b*) Base level: As in Exp. I, the mean base level at the time of CS onset for the first and last Threat CS was compared with that of the No-Threat CS to test for the influence of base level on the observed differences in CR magnitude to the two CSs. The *t* values for the 500- and 2,000-msec. interstimulus interval groups were .03 and .003, respectively, with 31 degrees of freedom.

Unwarned groups.—Figure 3 indicates the mean CR magnitude to the Threat and No-Threat CS at the 500 and 2,000 msec., CS-UCS intervals for those *S*s who were *not* forewarned of

---

[8] It will be noted that this procedure confounds stimulus duration with CS-UCS interval. An experiment conducted in this laboratory compared differential conditioning for a group with a 2,000-msec. stimulus and ISI vs. a 500-msec. stimulus with a 2,000-msec. ISI. No difference in differentiation was found in this study. It should also be noted that Kimmel (1967) did not obtain a significant decrement for a trace vs. delay effect on UCS magnitude for an ISI interval of 5 sec. for CSs lasting 1, 3, or 5 sec. It does not seem likely, therefore, that the confounding of ISI and stimulus duration is a significant variable in this experiment.

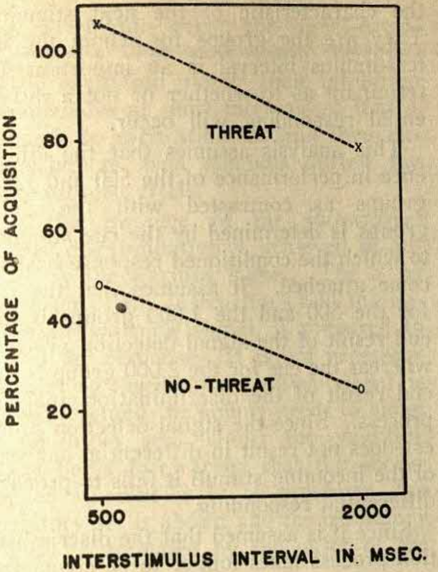


Fig. 2. Mean performance of the forewarned groups to the two stimuli at the 500- and 2,000-msec. intervals.

the nature of the CS prior to its onset. The main effect of Instructions was significant, $F\ (1, 28) = 20.64$, $p < .01$, as was the interaction of Instructions × Interstimulus Interval, $F\ (1, 28) = 10.45$, $p < .01$, indicating that the effect of instructions is dependent on the length of the interstimulus interval. A Newman-Keuls test of the means of Fig. 3 shows that the response to the Threat CS is significantly greater than the response to the No-Threat CS at the 2,000-msec. interval, $q_4\ (4, 28) = 7.79$, $p < .01$, but not at the 500-msec. interval. Since the response to the No-Threat CS does not change with the length of the interstimulus interval while the response to the Threat CS does, $q_2\ (2, 28) = 4.65$, $p < .01$, the superior discrimination at the long interstimulus interval may be attributed primarily to the increment in response to the Threat CS. The effect of Extinction Test was not significant, $F\ (3, 84) = 2.12$, $.05 < p < .10$.

Discussion

The dominant features of the data of the two experiments are as follows: (*a*) the uniformly powerful instructional effect under conditions of forewarning, (*b*) the dependence of the instructional effect on CS-UCS interval under conditions of no forewarning, and (*c*) the fact that if the conditions are favorable for the instructions to have an effect, the effect is stable and occurs relatively quickly.

In four of the six groups of the experiment the instructions produced differential performance to the two stimuli. These four instances included each group for whom forewarning was used and the nonforewarned group at the longest interstimulus interval. At the shorter intervals without forewarning the instructions were ineffective. This pattern of results clearly indicates that, although instructions are effective variables in determining the magnitude of the response, their effectiveness is limited by the operation of another variable associated with the interstimulus-interval value.

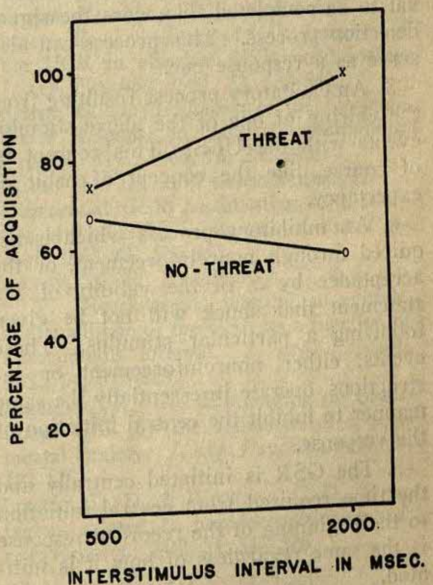


Fig. 3. Mean performance of the unwarned groups to the two stimuli at the 500- and 2,000-msec. intervals.

It has been suggested by Bridger and Mandel (1965) that failure to obtain differences following instructions may be attributed to disbelief by *S* that *E* will actually follow the new procedure described in the directions. It is indeed the case that some *S*s report that they do not believe *E*, but such an explanation will not handle the present data. In the second experiment *E* asked *S*s after the experiment whether they had believed him and most of them reported that they had. There were but two *S*s each in the 500 and 2,000 forewarned groups who expressed disbelief.

The results of the two experiments can be handled by the assumption that the following processes are involved:

1. A signal-detection process which indicates to the organism that a stimulus event has occurred (Wickens & Wickens, 1942). This process occurs rapidly and it does not include any indication about the modality of the stimulus. This process can serve as a cue for a response.
2. A discrimination process which differentiates one stimulus event from another. This process takes a longer interval to be completed than does the signal detection process. This process can also serve as a response cue.
3. An excitatory process resulting from the pairing of one of the above stimulus events with the UCS. This concept is, of course, like the concept of habit or expectancy.
4. An inhibitory process which is acquired through nonreinforcement, or the acceptance by *S* of the validity of *E*'s statement that shock will not be given following a particular stimulus. These events, either nonreinforcement or instructions, operate in essentially the same manner to inhibit the central initiation of the response.
5. The GSR is initiated centrally and the time required from central initiation to the beginning of the recorded response is the same regardless of how it is initiated.

These various concepts can account for the groups which were not forewarned of the characteristic of the next stimulus. They are the groups for whom the interstimulus interval is an important determinant as to whether or not a differential responding will occur.

This analysis assumes that the difference in performance of the 500 and 1,000 groups as contrasted with the 2,000 groups is determined by the cue stimulus to which the conditioned response has become attached. It assumes that the CS for the 500 and the 1,000 groups is the end result of the signal-detection process, whereas the cue for the 2,000 group is the end result of the discrimination analysis process. Since the signal-detection process does not result in differential analysis of the incoming stimuli it fails to produce differential responding.

Since it is assumed that the discrimination process takes longer to be completed than the signal-detection process, one would expect the latency of the CR to be greater for the 2,000 group than the 500 or 1,000. Unfortunately, latency data were not obtained in this experiment, but information on latency of a 500 and a 2,000 group are available from data collected by Hill (1967). Analysis of performance during extinction of a 500 and 2,000 group with 56 *S*s in each group obtained a mean latency of 1,812 msec. for the 500 group and 2,014 msec. for the 2,000 group. The difference is significant at the .002 level. It is pertinent to note that Prokasy, Fawcett, and Hall (1962) obtained no latency difference between a 5,000 and 1,000 group, though their 3,000 group is significantly slower. Thus the information obtained by combining these experiments indicates that the 500 and 1,000 groups are essentially alike, but the 2,000 group has a longer latency. Latency characteristics, therefore, parallel differences in the effect of the instructions.

In summary then, the 2,000 group behaves differentially to the two stimuli because the instructions establish an inhibitory set. In the instance of the shorter groups, however, differential responding to the two stimuli is not possible because the CR is attached to the same type of stimulus event, the signal-detection process. For some reason, the process of excitation rather than inhibition predominates at short intervals so that *S* responds to each stimulus, and then extinguishes to each stimulus during the nonreinforced trials.

In the instance of the forewarned groups, discrimination is achieved in advance by the warning signal. The *S*, as the result of the instructions and the appropriate warning signal, is prepared to inhibit the response. It is not necessary for him to discriminate which signal is being presented; he simply needs to adopt an inhibitory set. Under the condition where he is forewarned that the shock-implied stimulus is to be presented, the inhibitory set is absent and the excitatory process is activated by the onset of the stimulus.

An additional interpretation, stemming from Pavlov, is that for long delays the CS acquires inhibitory as well as excitatory characteristics (Kimmel, 1966). Under these circumstances, the mild effect of instructions could operate for the long interstimulus interval because some inhibition is already attached to it but not for the stimulus with the short interval.

It was mentioned in the introduction that there is some disagreement in the literature as to whether the instructions are effective immediately and completely or whether nonreinforced experience is also required to suppress responding. The results of the present experiments are somewhat ambiguous in this respect. In Exp. I the difference in the forewarned group on Trial 1 for the Threat and No-Threat stimulus was not significant, though it became significant by Trial 2. In the remaining instances where instructions had an effect, the effect was significant from the first trial on and tended to increase, though not significantly on successive trials. On the basis of these findings it seems appropriate to conclude that, given the proper interstimulus interval, instructions have a powerful immediate effect though the effect is not complete and requires some nonreinforced trials to become complete.

The overall trend of the results of the two experiments indicates that the behavior following instructions does not simply reflect *S*'s belief or disbelief in the validity of the instructions, for the behavior is also very clearly a function of the objective relationship between CS and UCS onset and the amount of foreknowledge of the nature of the CS to occur.

REFERENCES

BRIDGER, W. H., & MANDEL, I. J. Abolition of the PRE by instructions in GSR conditioning. *J. exp. Psychol.*, 1965, **69**, 476–482.

GRINGS, W. W., & LOCKHART, R. A. Effects of "anxiety-lessening" instructions and differential set development on the extinction of GSR. *J. exp. Psychol.*, 1963, **66**, 292–299.

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

HILL, F. A. The effects of instructions and subject's need for approval on the GSR. *J. exp. Psychol.*, 1967, **73**, 461–467.

KIMMEL, E. Judgments of UCS intensity of the UCR in classical GSR conditioning. *J. exp. Psychol.*, 1967, **73**, 532–543.

KIMMEL, H. D. Inhibition of the Unconditioned Response in classical conditioning. *Psychol. Rev.*, 1966, **73**, 232–240.

KIMMEL, H. D., & PENNYPACKER, H. S. Differential GSR conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1963, **65**, 559–563.

PROKASY, W. F., FAWCETT, J. T., & HALL, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of interstimulus interval. *J. exp. Psychol.*, 1962, **64**, 513–518.

SILVERMAN, R. E. Eliminating a conditioned GSR by the reduction of experimental anxiety. *J. exp. Psychol.*, 1960, **59**, 122–125.

WICKENS, D. D., & WICKENS, C. D. Some factors related to pseudoconditioning. *J. exp. Psychol.*, 1942, **31**, 518–526.

(Received July 13, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 370–377

PRIMING OF ASSOCIATION TEST RESPONSES BY DIFFERENTIAL VERBAL CONTEXTS[1]

SYDNEY JOELSON SEGAL

Research Center for Mental Health, New York University

The associative frequency of words was measured following their activation either intentionally in a learning task, or incidentally, embedded in an analogy test. The words used were homographs, whose meaning was systematically changed by varying the embedding context of the analogy test. 3 types of context were used: 1 identical to the context provided by a stimulus word on the association test, 1 congruent, and 1 incongruent. A strong "priming effect," i.e., raising of the associative frequency, was obtained after identical contexts, a weaker but significant effect after congruent contexts or learning, and no effect at all after incongruent contexts. 3 experiments are reported with similar findings: in 1, the use of a continued association test demonstrated that the primed word always occurred in the 1st response position, and that priming declined over time, but at different rates depending on the pretask.

Selected associates on an association test occur more frequently when they have just been presented to *S*s in a different pretask. This phenomenon, called "recency" when first reported by Storms (1956, 1958) and described as "priming" in our work (Grand & Segal, 1966; Segal, 1966, in press) is a common experience familiar to speakers and writers who often find that the most recently used word comes to mind in trying to express a related idea. The findings of Storms have been confirmed (Cofer, 1956; Segal, 1966, in press) using a variety of different pretasks. Storms (1958) and Russell (1955), in reporting on Storms' work, suggested that this was a performance effect, dependent on the principle of recency, which temporarily increased the response strength. Recency implies that the mere recent exposure of a list of responses would alter their associative frequency; the present writer's unwillingness to accept this premise led to the adoption of the term "priming."

The studies reported thus far, however, have demonstrated only that priming occurs over a variety of different pretasks, different word lists, and with different associative relationships. Indeed, over most of the tasks, the response frequency of the critical words has generally increased by about 5% to 12% over the norm frequency, that is, an increase of about 1 out of 14 critical words. The generality and smallness of this effect can readily be subsumed under the assumption of a "performance" effect.

In the present experiments, the pretask was selected to provide varied semantic contexts, in the expectation that this would lead to better differentiated priming effects. It has been

---

[1] This study was supported in part by a United States Public Health Service Predoctoral Fellowship from the National Institute of Mental Health, Grant No. MH-13, 324 and by National Institute of Mental Health Research Grant No. MH-06733.

Grateful acknowledgment is made of the help rendered by Charles N. Cofer, who was a gifted and efficient sponsor of the thesis, and who has advised on numerous drafts of the present paper. Thanks also to Donald P. Spence and Morris Eagle, who read and criticized different drafts of the paper.

shown that S's profession (Foley & Macmillan, 1943) or sex (Goodenough, 1946), the verbal context (Howes & Osgood, 1954), even the condition of administration (Siipola, Walker, & Kolb, 1955), may affect associations in unpredictable ways. Similarly, any change in the semantic context of the pretask in which priming is effected should alter the subsequent associative availability of the word.

In order to maximize the effect, the present experiment resorted to the use of homographs, words with two distinct meanings but the same spelling. Varying the context of the pretask may facilitate or inhibit the association, depending on how the meaning of the word implied in the priming context relates to the meaning of the word when it is elicited by the cue stimulus on the association test.

EXPERIMENT I

Method

Subjects.—Three hundred thirty-seven experimental and 70 control *S*s participated in this study. All were undergraduate students enrolled in elementary and intermediate psychology classes at Brooklyn College.[2] All were tested as groups, during a class period, replicating the procedure used by Storms in his original report.

Materials and procedure.—The *S*s were handed a booklet containing two pages. First, they were asked to solve an analogy test of seven items, printed on the first page (cf. Table 1). Fourteen critical words were embedded in these seven items, two in each. The *S*s were given 4 min. to solve the analogies, and all were finished within this time. Immediately after, they were asked to do "a different task, an association test." The second page in the booklet contained a 35-item discrete word-association test, and *S*s wrote in their associations. Fourteen of the stimuli were cues, known to elicit the 14 critical words, 14 were cues for 14 matched-control associations, and 7 were buffer items. All *S*s completed the association test within 2 min. The measure of priming was the increased response frequency of the critical words.

Analogy test.—The critical words were all homographs, that is, words with two or more meanings and a single spelling. The cue stimulus on the association test mediated only one of the possible meanings, "river," for example could elicit "bank" as an associate, but not the (First National) Bank. Each homograph appeared in one analogy test in a context identical to the cue stimulus ("river is to bank"), in one in a context which mediated a congruent meaning, and in one in a context which mediated an incongruent meaning. The three analogy tests, and the 14 cue stimuli, are listed in Table 1. Each experimental *S* took one analogy test, and therefore worked with all 14 words, 4 in one context, 5 in each of the other two. The data from the 240 *S*s were especially combined, and reported as if they came from 80 "incongruent" *S*s, 80 "congruent," and 80 from an "identical" condition.

Learning.—To compare context priming more readily with previous work, two intentional learning conditions were also introduced. In one ($N = 59$), the 14 homographs were presented to *S*s for learning: Each word was called out successively, while an index card was held up on which the word was written in inch-high letters in India ink. Afterward, *S*s took the association test. In the other learning task ($N = 38$), the 49 words included in the seven analogies were presented without connectives, as a list of words for learning. The *S*s were given 4 min. to try to learn them. The association test followed. After the association test, both groups were tested for recall.

Normative group.—Two classes of students ($N = 38$, 32) took the same association test used in the other conditions, but with no pretask; from the responses of these *S*s, it was possible to estimate the norm frequency of occurrence of the homographs to the cue stimuli, and to select 14 control associates, matched individually to the critical associates for frequency of occurrence.

Results

Measures of priming.—Any variability among different *S*s or different classes of *S*s due to intrinsic verbal characteristics, or set and intention, could be assumed to affect the critical

---

[2] The kindness of the instructors who permitted use of their class time is gratefully acknowledged: Evelyn Raskin, Marvin Zuckerman, Austin Wood, William Ittleson, Magda Denes, and Pearl Meissner.

TABLE 1

ANALOGIES TESTS AND WORDS USED IN EXP. I AND II

Type of Context[a]		
First	Second	
		Form a:
iden.	cong.	1. PLANT[b] is to GREEN as BANK[b] is to a. Stream b. Wet c. Sandy d. Lake
iden.	cong.	2. SICKNESS is to WELL[b] as TOP[b] is to a. Below b. Bottom c. Valley d. Peak
inc.	inc.	3. SCALE[b] is to FISH as DOWN[b] is to a. Pear b. Peach c. Whale d. Banana
iden.	cong.	4. CITY is to STATE[b] as BULB[b] is to a. Lamp b. Brilliance c. Chandelier d. Socket
iden.	inc.	5. KING is to RULER[b] as ORGAN[b] is to a. Piano b. Percussion c. Composer d. Violin
iden.	inc.	6. HARD is to ROCK[b] as SOUND[b] is to a. Safe b. Secure c. Whole d. Dollar
cong.	inc.	7. TENNIS is to NET* as BOWL[b] is to a. Alley b. Strike c. Ball d. Tenpin
		Form b:
iden.	inc.	1. BULB[b] is to LIGHT as WELL[b] is to a. Water b. Wish c. Pennies d. Bucket
iden.	cong.	2. TOP[b] is to MOUNTAIN as DOWN[b] is to a. Under b. Low c. Plain d. Valley
iden.	cong.	3. ORGAN[b] is to STOMACH as STATE[b] is to a. County b. Province c. Manhattan d. Connecticut
iden.	cong.	4. RIVER is to BANK[b] as SOUND[b] is to a. Noise b. Listen c. Horn d. Drum
inc.	inc.	5. GROSS is to NET[b] as PLANT[b] is to a. Industry b. Manufacture c. Production d. Machinery
inc.	cong.	6. RULER[b] is to INCH as SCALE[b] is to a. Balance b. Weight c. Honest d. Pound
iden.	cong.	7. BOWL[b] is to FRUIT as ROCK[b] is to a. Surface b. Solid c. Barnacles d. Wall
		Form c:
inc.	inc.	1. SPIN is to TOP[b] as ROCK[b] is to a. Baby b. Crib c. Roll d. Cradle
cong.	inc.	2. BOWL[b] is to CONTAINER as BANK[b] is to a. Money b. Loans c. Institution d. Investments
iden.	cong.	3. NET[b] is to BUTTERFLY as ORGAN[b] is to a. Tissue b. Nourishment c. Gland d. Secretion
iden.	cong.	4. DOWN[b] is to DEEP as WELL[b] is to a. Health b. Disease c. Healthy d. Happy
inc.	cong.	5. BULB[b] is to TULIP as PLANT[b] is to a. Chlorophyll b. Rhododendron c. Leaves d. Blossom
iden.	inc.	6. SOUND[b] is to MUSIC as STATE[b] is to a. Fact b. Lecture c. Poem d. Gossip
iden.	cong.	7. JUSTICE is to SCALE[b] as RULER[b] is to a. Emperor b. Queen c. Sceptre d. Subjects

[b] Critical Words	Cue Stimulus on Association Test
Plant	GREEN
Well	SICKNESS
Scale	JUSTICE
State	CITY
Ruler	KING
Rock	HARD
Net	BUTTERFLY
Bulb	LIGHT
Top	MOUNTAIN
Organ	STOMACH
Bank	RIVER
Bowl	FRUIT
Down	DEEP
Sound	MUSIC

---

[a] Type of context in first half and second half of each item, iden. = identical, inc. = incongruent, cong. = congruent.

TABLE 2

Priming Effect in the Experimental Conditions; Exp. I

Cond.	N	Mean No. Critical Responses	Mean No. Control Responses	Mean Responses Minus Control Responses
Norms	70	.77	.76	.01
Learning (Controls)				
14 word	50	1.46	.92	.54
49 word	38	1.16	.63	.53
Experimental	240			
Identical	(80)	3.45	.64	2.81
Congruent	(80)	1.15	.64	.51
Incongruent	(80)	.80	.64	.16

and the control associates equally. The excess of critical over control associates, therefore, is a specific measure of priming. This measure was used for the learning groups. For the context groups, a comparable figure was obtained. The 240 *S*s were grouped into 80 triplets: Each triplet included three *S*s who took the 3 forms of the analogy test and who obtained the same score on the 14 control words.[3] By summing across the three association tests, three scores were obtained representing priming to 14 words after exposure in (*a*) a congruent, (*b*) an incongruent, and (*c*) an identical context. The three scores could then be compared directly to the learning data.

Table 2 lists the mean number of critical responses and control responses for each of the 5 conditions and the normative group. The critical and control responses are, by definition, equal in the norms; the control responses vary only slightly among the experimental groups, while the critical responses are higher. The mean difference scores are reported in the last column, and show that there was virtually no priming after incongruent priming, that the congruent context yielded as much priming as the learning conditions; and the identical context produced a large priming effect. An analysis of variance, based on these difference scores, was significant, F $(5, 401) = 37.0, p < .001$. It was also significant ($p < .01$) excluding the strong identical condition, however, Duncan *t* tests, based on the error variance, were only significant for the identical condition, which differed from all others.

A second analysis of variance was also performed, duplicating the procedure of Storms, in his initial report. For each experimental condition, the percentage of responses to each of the 14 critical words was recorded; after arc-sin conversions, an analysis of variance was performed. It was significant, $F\ (5, 78) = 10.8,\ p < .01$.

Experiment II

Experiment II was carried out to repeat the conditions of Exp. I, and also answer (*a*) what happens to the priming effect over time; and (*b*) where in the response hierarchy the primed word occurs.

Method

All the experimental variations of Exp. I were replicated; the difference was that, instead of using the 35-item discrete free word-association test, *S*s were given a continued association test. They had a booklet, containing the analogy test and 14 pages, each

TABLE 3

Priming Effect in the Experimental Conditions, Continued Word Associations; Exp. II

Cond.	N	Mean No. Responses	Mean Responses for Control Words	Mean Responses Minus Control Word Responses
Norms	78	3.02	3.00	.02
Learning				
14 word	42	4.09	2.35	1.74
49 word	45	3.47	2.25	1.22
Experimental	132			
Identical	(44)	6.97	2.51	4.46
Congruent	(44)	3.70	2.51	1.19
Incongruent	(44)	2.67	2.51	.16

---

[3] In a few instances, this requirement could not be met perfectly.

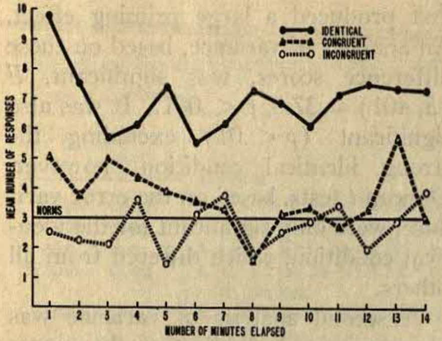


FIG. 1. Effect of priming over 14 min. of continued association for the context conditions.

listing only one cue stimulus. After the analogy test (or the learning task), they responded to each word for 1 min., giving as many associations as possible. Order of presentation of the cue stimuli was randomized. The cue stimuli and critical associates were identical with those of Exp. I; the control associates were chosen from the norms. Each control associate occurred in response to a cue, but never to the same cue as its matched critical associate.

Results

The data, presented in Table 3 closely parallel the results of Exp. I. The three context groups again differ; even most of the nonsignificant differences were paralleled. Two analyses of variance, comparable to those of Exp. I, were both significant: the Storm's replication, based on the percentage priming in the 14 words, F (5, 78) = 8.25, $p < .05$; the other, based on the difference between critical and control associative frequency, F (5, 291) = 34.8, $p < .001$. A Duncan t test, based on the analysis of variance of differences, was only significant for the identical condition; although again, the analysis was still significant ($p < .01$) when the identical condition was excluded.

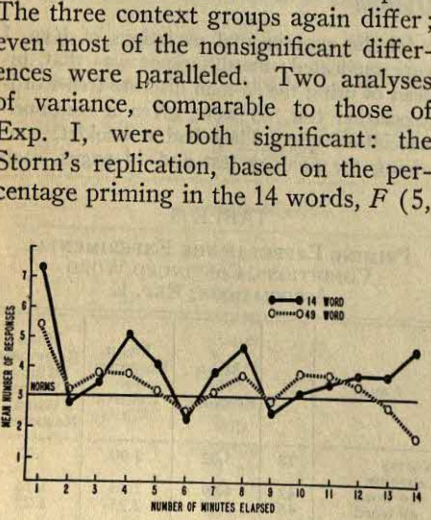


FIG. 2. Effect of priming over 14 min. of continued association for the learning conditions.

Priming over time.—As $\bar{S}$ responded to each word for 1 min. and the order of words was randomized, it was possible to trace the effect of priming over 14 min. Figures 1 and 2 illustrate these effects. For the two learning groups, most of the priming effect occurred in the first minute. For the congruent and incongruent groups, there was more moderate priming extending over 4 to 5 min. Identical priming remained effective throughout the 14 min. Table 4 demonstrates that, except for identical priming, the entire priming effect occurred within the first 5 min. When the first and last 5 min. are compared, there is a significant *increase* for the incongruent condition, a nonsignificant decline for the norms and the 49 word learning group, and a significant decline for all others.

Hierarchic occurrence of primed word.—It has been demonstrated that

TABLE 4

PRIMING EFFECT OVER TIME; FIRST 5 MIN. OF CONTINUED ASSOCIATION TEST COMPARED WITH LAST 5 MIN.

Group	Mean Number of Responses			
	First 5 Min.	Last 5 Min.	t	p
Control	.248	.223	2.10	
Incongruent Context	.182	.234	2.77	.01
Congruent Context	.327	.268	3.29	.01
Identical Context	.523	.432	5.80	.01
Learning, 14 word	.343	.281	3.25	.01
Learning, 49 word	.262	.227	1.87	

TABLE 5

LOCATION OF THE PRIMED RESPONSE WITHIN THE ASSOCIATING MINUTE; CONTINUED ASSOCIATION EXP. II

Response Position	Frequency of Occurrence of Primed Responses for Each Condition				
	Norms	Learning (14 words)	Incongruent	Congruent	Identical
First	.60	1.28	.83	1.29	3.29
Second, third, fourth, fifth, (comb.)	.45	.37	.29	.37	.71
Sixth or later (comb.)	.20	.26	.15	.14	.16

in continued associations, the first associate is usually a high frequency word, the second less common, and so on (Cofer, 1958; Rosen & Russell, 1957). If the critical word is strengthened within an existing hierarchy, it should occur earlier than usual. Table 5 demonstrates that this is indeed the case. Almost the entire priming effect is located in the first response position. That is, if priming occurs, the critical word generally is the first associate elicited, occasionally the second; but if it does not occur first or second, there is no priming at all.

EXPERIMENT III

Experiment III was a replication of the congruent and incongruent conditions of the previous experiments. As the difference between these two conditions was generally small and not always significant, a further replication seemed desirable, especially as this difference was the most crucial. The conditions were therefore repeated here, using a new and larger group of words.

Method

Twenty homographs were selected, and two analogy tests were prepared of 10 items each. Each test contained all 20 homographs, half in a congruent context, half, in an incongruent. A 50-item discrete word-association test was used, including the 20 cue stimuli, 20 cues for the control associates, and 10 buffer items. The data were summed across both tests, as in the previous experiments, resulting in two "groups," ($N = 28, 28$) an incongruent and a congruent. A small control group ($N = 28$) was also used.

Results

The analysis of variance based on difference scores failed to meet significance, $F\ (2, 81) = .98$. The analysis of variance based on the percentage of priming of the 20 words, was significant, $F\ (2, 59) = 3.5$, $p < .05$. The difference between congruent and incongruent priming, however, was in the same direction and of the same size as in the two earlier experiments.

Discussion

The main finding of these experiments is that the availability of a word is closely dependent on its meaning. Without varying the nature of the pretask, and varying only the verbal context in which the word occurs, priming can be facilitated or inhibited. The identical condition is a special case, more similar to paired associate learning than to "priming" as used here. The congruent and incongruent context conditions, however, illustrate the differentiating effects of a semantic context. It appears that a word, that has been activated, is associatively available only in a related context.

TABLE 6

PRIMING EFFECT IN THE EXPERIMENTAL CONDITIONS; EXP. III

Cond.	N	Mean No. Responses	Mean Responses Control Words	Mean Responses Minus Control Word Responses
Norms	28	1.60	1.60	.0
Incongruent	28	2.00	1.82	.18
Congruent	28	2.68	1.82	.86

Another way of describing this is to say that availability is not sufficient without relevant differentiation (Saltz, 1961): the mere occurrence of a word in a pretask is not sufficient. If the word occurs in an incongruent context, it is semantically predifferentiated into a nonrelevant category. In this condition, recency was inoperative: the word was exposed in a pretask, but no priming occurred. Indeed, when the associations were measured by continued associating, a condition which favors meaningful associations at the expense of verbal-motor habits (Cofer, 1958) there is evidence of a significant decline in the frequency of occurrence of the critical word. That is, in this condition, the tendency to elicit the "primed" response is inhibited. After the word has been predifferentiated in a relevant context, the priming effect is similar in degree to the effect after learning, but it seems somewhat more persistent over time.

The learning tasks show a priming effect which can readily be described as a "performance" effect, due to recency. The main priming effect occurred in the first minute, and it appeared to affect the first response position without any change in the later responses. As Storms used a learning pretask, it is quite appropriate that he referred to the phenomenon as recency.

As noted, the identical context condition is really an example of paired-associate learning: Both the eliciting cue stimulus and the primed associate were presented in the pretask, joined in a meaningful association; hence there was priming or learning of the pair. In the association test, one member of the pair was given as a cue to measure how often it elicited the other member of the pair. The increase of 3 or 4 associates in this condition, its durability over the 14 min. of continued association, and its occurrence both in the first and second response positions, may be the operational maximum for priming. If 4.47 responses, or 32%, is the maximum, then 68% of the responses were unaffected. This gives greater weight to the modest priming effects reported in the other conditions; it also demonstrates that the natural associative hierarchy shows some degree of resistance to momentary effects of recent experience.

Thus, these experiments do not disprove Russell and Storms' hypothesis about a recency effect, but suggest that such a conceptualization, if not inaccurate, is slightly misleading. Recency is effective only if the semantic context is relevant; the performance is changed only if there has been appropriate predifferentiation of the critical associate. In a simple task, such as list learning, the effect seems largely due to recency, and can be considered a performance effect. However, in actual daily practice, words are generally activated within a context, usually the context of connected discourse. These data suggest that the "recency" effects are relevant only in the artificial lists of the laboratory; once words are activated within semantic structures, recency alone is not sufficient, but there must also have been relevant predifferentiation.

REFERENCES

Cofer, C. N. Learning, retention, and recovery of meaningful material. An Exploration of the "Recency" effect in word association. Technical Report No. 5, 1956, University of Maryland. Contract Nonr-595 (04), Office of Naval Research.

Cofer, C. N. Comparison of word associations obtained by the method of discrete single word and continued association. *Psychol. Rep.*, 1958, **4**, 507–510.

Foley, J. P., Jr., & Macmillan, Z. L. Mediated generalization and the interpretation of verbal behavior. V. 'Free association' as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.

Goodenough, F. Sematic choice and personality structure. *Science,* 1946, **104**, 451–456.

Grand, S., & Segal, S. Recovery in the absence of recall: An investigation of color-word interference. *J. exp. Psychol.*, 1966, **72**, 138–144.

Howes, D., & Osgood, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.,* 1954, **67**, 241–258.

Rosen, E., & Russell, W. A. Frequency characteristics of successive word associations. *Amer. J. Psychol.,* 1957, **70**, 120–122.

Russell, W. A. Bidirectional effects in word association. Paper read at Minnesota Conference on Associative Processes in Verbal Behavior, Minneapolis, 1955.

Saltz, E. Response Pretraining: Differentiation or availability. *J. exp. Psychol.,* 1961, **62**, 583–587.

Segal, S. Priming compared to recall: Following multiple exposures and delay. *Psychol. Rep.,* 1966, **18**, 615–620.

Segal, S. The priming of association test responses: Generalizing the phenomenon. *J. Verbal Learn. Verbal Behav.,* in press.

Siipola, E., Walker, W. N., & Kolb, D. Task attitudes in word association, projective and non-projective. *J. Pers.,* 1955, **23**, 441–459.

Storms, L. H. Backward association in verbally mediated learning. Unpublished doctoral dissertation, University of Minnesota, 1956.

Storms, L. H. Apparent backward associations: A situational effect. *J. exp. Psychol.,* 1958, **55**, 390–395.

(Received August 1, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 378–382

SIZE MATCHING AS A FUNCTION OF INSTRUCTIONS IN A NATURALISTIC ENVIRONMENT[1]

H. W. LEIBOWITZ AND LEWIS O. HARVEY, JR.

Pennsylvania State University

The sizes of human being test objects were matched under 3 sets of instructions in an outdoor environment at distances from 340 to 1,680 ft. Matches obtained under "objective" instructions corresponded to the true size of the test object at all distances. With "apparent" instructions, matches ranged from 36% to 20% of test-object size, decreasing as a function of distance. Under "retinal" instructions, matches were from 18% to 9% of true size. It is concluded that there is a family of functions relating matched size to distance, with instructions as an important parameter. Theoretical implications of the observed effect of instructions and the differences between perceived phenomenal size and matched size are discussed.

The phenomenon of size constancy refers to the observation that the matched size of an object tends to remain invariant in spite of reduction in retinal image size as distance is increased. A number of studies indicate that matched size tends to remain constant at distances up to 100 ft. or more (Gibson, 1950; Gilinsky, 1955; Graham, 1965; Holway & Boring, 1941; Smith, 1953). In some cases, there is an overestimation of the object's true size, referred to as overconstancy (Carlson, 1960, 1961; Wohlwill, 1963).

Following early investigators (Holaday, 1933; Holway & Boring, 1941; Martius, 1889), most size-matching studies in the literature have employed test objects such as circles, triangles, or sticks, which have no definite size and can, logically, exhibit a wide range of dimensions. The question may be raised whether the results of such studies are representative of perception outside the laboratory, since in our daily lives one views familiar objects which, from past experience, are known to have either a definite size or a restricted range of possible sizes. One basis for these doubts is that in the phenomenal perception of the world, objects do, in fact, appear smaller the farther they are from us (Helmholtz, 1962, p. 282). At the same time we know that the physical size of these objects remains the same. Indeed, one of the classical monocular cues for depth, apparent size (Boring, 1942), which assumes that distance may be inferred from the decreasing phenomenal size of familiar objects, is logically inconsistent with the concept of size constancy which states that phenomenal size remains invariant with distance. If the apparent size, i.e., phenomenal size, of familiar objects is invariant with distance, apparent size could not, at the same time, serve as a cue for distance.

In order to clarify these uncertainties in the present literature, the present experiment was designed with the following objectives: (*a*) to obtain size matches for familiar objects having a fixed and known size, viewed in a naturalistic environment; (*b*) to obtain matches at distances greater than those usually employed in laboratory studies,

---

[1] Supported by Grant MH 08061 from the National Institute of Mental Health. The authors gratefully acknowledge the assistance of Warren Selekman, Sharon Toffey Shepela, and Stephen Pollard.

e.g., beyond 100 ft.; (*c*) in view of the marked effect of instructions for test objects of indefinite size (Carlson, 1960, 1962; Carlson & Tassone, 1962; Gilinsky, 1955; Holaday, 1933; Jenkin & Hyman, 1959), to determine the extent to which this variable can influence *S*'s matches for a familiar test object having a known and fixed size.

METHOD

Subjects.—The *S*s were 43 male and female volunteers from elementary psychology classes, who were unaware of the purpose or nature of the experiment. All *S*s had normal or corrected vision and observed binocularly.

Apparatus and procedure.—The test objects were human beings who stood in the middle of a mall on the university campus at distances of 340, 680, 1,020, 1,360, or 1,680 ft. from *S*. The mall is flanked on either side by rows of trees, paths, fences, and buildings. At all times, there were a large number of people moving about in all directions on the mall. The *S*s stood on the steps of the library facing the mall and viewed the comparison stimulus, a wooden rod (surveyor's leveling rod) 2 in. wide, which protruded vertically from behind a wall 51 ft. from *S* and 90° to his left. The base of the visible portion of the rod was level with *S*'s feet.

The experiment was divided into two parts. In Part 1, the test stimulus was either a female (Fall 1964) or a male (Summer 1965) graduate student, each 5.8 ft. in height, who stood at the above five distances. Ten *S*s matched the height of the female test object and 10 different *S*s matched the height of the male test object. At the beginning of the experiment, *S*s were introduced to the human test object. The *S*s were instructed to judge which *appeared* larger, the human test object or the comparison rod.

In Part 2, two sets of instructions were employed: "objective," which instructed *S* to match the physical height of the test objects, irrespective of distance; and "retinal," which instructed *S* to match in terms of the visual angle. These instructions are identical to those used by Gilinsky (1955), except for minor changes to adapt them to the present experimental procedure. The *S*s were questioned to insure understanding of the instructions. Thirteen *S*s made size judgments on the same human test object, 5.8 ft. high, who moved to the five test distances. In order to eliminate the possibility that *S*'s judgments were influenced by the knowledge that the physical height of the one test person does not change, 10 additional *S*s made size judgments for five different human test objects stationed at the five different test distances. The heights of these test objects ranged from 5.7 ft. to 6.1 ft., with a mean height of 5.9 ft. Half of the *S*s received the objective instructions first and the other half received the retinal instructions first.

In both parts of the experiment, the order of presentation of distances was determined by a Latin-square design. The method of limits was employed. No time limit was placed on *S*, who completed his observations within approximately 30 min. All observations were made in good weather.

RESULTS

Because the test objects varied in size, Brunswik ratios (BR) are used to represent the extent to which *S*'s matches are in accordance with size constancy (Brunswik, 1933). A BR of 1.0 indicates agreement with a prediction based on the law of size constancy, i.e., correct size matches, while a BR of 0 indicates agreement with a prediction based on the law of the retinal image. Because the use of two different human test objects in Part 1 produced no significant differences in size matches obtained under "apparent" instructions, and as there were no significant differences between data from the individual test object and from the five different test objects under the objective and retinal instructions in Part 2, the combined BRs for each of the sets of instructions at each of the five test distances are plotted in Fig. 1.

It is clear that instructions have a profound effect on matched size. With the objective instructions, *S*s were able to match the true size of the human test objects quite accurately. The errors were all in the direction of over-

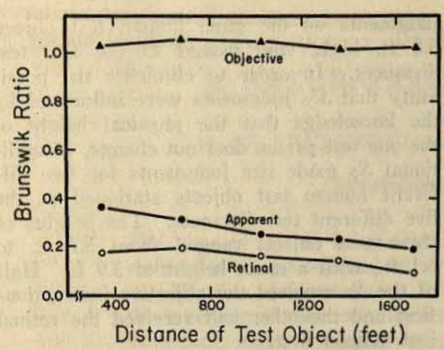


FIG. 1. Brunswik ratio (percentage of constancy) as a function of distance under three sets of instructions. (Test objects were human beings viewed in an outdoor environment.)

estimation, ranging from 2% to 6% of the correct size. This overestimation is not nearly as large as Gilinsky (1955) reported for her *S*s, and does not increase with distance. However, with the apparent instructions, similar to those employed in many studies reporting agreement with the law of size constancy (Holway & Boring, 1941; Zeigler & Leibowitz, 1957), the present data indicate a marked lack of constancy. There were no BRs above .36, and the BRs decreased with distance. With the retinal instructions, the range of BRs was even lower, from .18 to .09. In other words, even with explicit instructions emphasizing the visual angle, *S*s are *not* able to make accurate retinal size matches, but consistently overestimate the visual angle subtended by the test object. These retinal image matches are much lower at distances closer than 400 ft. than those of Gilinsky, who reported BRs as high as .5.

There was no overlap between either the objective and apparent distributions or the objective and retinal distributions at any distance. Table 1 presents a summary of a two-way analysis of variance testing the effect of apparent vs. retinal instructions and the effect of distance on matched size. The table shows that apparent instructions produced significantly higher BRs than did retinal instructions. Also, matched size decreased significantly as a function of distance.

DISCUSSION

The outstanding feature of the present data is the marked effect of instructions on matched size. Whereas the BRs for the objective instructions are approximately 1.0, i.e., in agreement with the law of size constancy, the ratios for apparent instructions are less than half this value while those for retinal instructions approach a prediction based on the law of the visual angle. Thus, the variable of instructions is seen to produce changes in size-matching behavior which encompass a range of values between a prediction based on size constancy to a level which approaches a retinal-image match. This observation has been previously well-established in the literature for unfamiliar test objects. The unique feature of the present data is that this effect is here seen to be demonstrated for familiar objects having a known, fixed, and definite size.

A major difference between the results of the present study and the majority of previous experiments is evidenced by the matches obtained under apparent instructions. These instructions, utilized in a laboratory situation with unfamiliar test

TABLE 1

ANALYSIS OF VARIANCE SUMMARY TABLE: APPARENT INSTRUCTIONS VS. RETINAL INSTRUCTIONS AND DISTANCE

Source	*df*	*F*
Apparent vs. Retinal	1	5.94*
Error (Instructions)	41	
Distance	4	16.14**
Instructions × Distance	4	2.52*
Error (Distance)	164	
Total	214	

* $p < .05$.
** $p < .01$.

objects of indefinite size, usually produce matches which approximate the law of size constancy (see review by Epstein, Park, & Casey, 1961). However, the same instructions, with respect to the familiar test objects and naturalistic setting of the present study, produce size matches which not only are less than half the value predicted by the law of size constancy, but also decrease as a function of distance. It should also be noted that the matches obtained under "objective" and "retinal" instructions are consistently lower than those obtained by Gilinsky using the same instructions in an outdoor environment over a comparable range of distances. Since her *S*s made matches for a wooden triangle, this would suggest that the nature of the test object is the basis for the lower matches obtained with the familiar test objects in the present experiment.

An implication of the present results is that the matched size of objects obtained in experimental studies does not necessarily correspond to their phenomenal size. Rather, matched size is seen to be strongly influenced by instructions and may assume a wide range of values depending upon the particular experimental conditions. There is, in fact, a family of functions relating matched size to distance, some of which exhibit characteristics incompatible with perceived size. For example, size matches obtained under "objective" instructions are invariant with distance, whereas the phenomenal size of objects decreases with distance. It may not be possible, methodologically, to quantify phenomenal size since experimentally we can measure only size-matching behavior. It is felt, however, that the data obtained under "apparent" instructions in the present context represent a first approximation to phenomenal size since this function is similar to everyday experience.

It is clear that regardless of the magnitude of size matches made under a given set of conditions, information regarding the true sizes of unfamiliar as well as familiar test objects is available to *S*, since veridical size matches can and frequently are made. (Exceptions would be observation of unfamiliar objects under reduction conditions, astronomical bodies, and extremely long observation distances.) This raises the question as to why size-matching behavior can exhibit so many functional relations and why matched size and phenomenal size may differ. To resolve this discrepancy, the following hypothesis is offered: In normal, everyday perception, one does not ordinarily emphasize the sizes of objects. Rather, the observer is concerned with many variables, sensory as well as nonsensory, of which size is only one, and in many cases, a relatively insignificant aspect. However, in the experimental situation, attention is concentrated on the size dimension with the result that *S* brings into play all available information regarding the size of the test objects, much of which would be ignored in casual observation. This information may be supplied by a number of mechanisms known to influence perceived size, such as familiarity with the particular test-object or class of test objects in question (Bolles & Bailey, 1956), data supplied by context cues (Boring, 1946), and, for short-observation distances, information from oculomotor feedback (Harvey & Leibowitz, 1967; Leibowitz & Moore, 1966). In a size-matching task, it is assumed that constancy is the rule, not only by virtue of the fact that information regarding the true size of the object is available, but also because special attention is drawn to the size dimension by the very nature of the experimental situation. Instructions are assumed to perform a *selective* function, determining whether the size information available to *S* will be manifested in size-matching behavior. In the absence of explicit size-matching instructions, all available information does not come into play, with the result that phenomenal size decreases with distance, one consequence of which is to permit the observer to utilize apparent size as a distance cue. However, this decrease is not as marked as would be predicted from the dimensions of the retinal image. Phenomenal size represents a compromise between knowledge of the true size of the test object and the diminishing size of the retinal image.

REFERENCES

Bolles, R. C., & Bailey, D. E. Importance of object recognition in size constancy. *J. exp. Psychol.,* 1956, **51**, 222–225.

Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts, 1942.

Boring, E. G. The perception of objects. *Amer. J. Physics,* 1946, **14**, 99–107.

Brunswik, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung und deren quantitative Bestimmung. *Arch. für die gesamte Psychol.,* 1933, **88**, 377–418.

Carlson, V. R. Overestimation in size constancy judgments. *Amer. J. Psychol.,* 1960, **73**, 199–213.

Carlson, V. R. Effects of sleep-deprivation and chlorpromazine on size constancy. *Amer. J. Psychol.,* 1961, **74**, 552–560.

Carlson, V. R. Size constancy judgments and perceptual compromise. *J. exp. Psychol.,* 1962, **63**, 68–73.

Carlson, V. R., & Tassone, E. P. A verbal measure of the perspective attitude. *Amer. J. Psychol.,* 1962, **75**, 644–647.

Epstein, W., Park, J., & Casey, A. The current status of the size-distance hypothesis. *Psychol. Bull.,* 1961, **58**, 491–514.

Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

Gilinsky, A. S. The effect of attitude upon the perception of size. *Amer. J. Psychol.,* 1955, **68**, 173–192.

Graham, C. H. Visual space perception. In C. H. Graham (Ed.), *Vision and visual perception.* New York: Wiley, 1965. Pp. 504–547.

Jenkin, N., & Hyman, R. Attitude and distance-estimation as variables in size matching. *Amer. J. Psychol.,* 1959, **72**, 68–76.

Harvey, L. O., Jr., & Leibowitz, H. W. Effects of exposure duration, cue reduction, and temporary monocularity on size matching at short distances. *J. Opt. Soc. Amer.,* 1967, **57**, 249–253.

Holaday, B. E. Die Grössenkonstanz der Sehdinge bei variation der inneren und äusseren Wahrnehmungsbedingungen. *Arch. für die gesamte Psychol.,* 1933, **88**, 419–486.

Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.,* 1941, **54**, 21–37.

Helmholtz, H. V. *Physiological optics.* Vol. III. New York: Dover, 1962.

Leibowitz, H., & Moore, D. Role of changes in accommodation and convergence in the perception of size. *J. Opt. Soc. Amer.,* 1966, **56**, 1120–1123.

Martius, G. Ueber die scheinbare Grösse der Gegenstände und ihre Beziehung zur Grösse der Netzhautbilder. *Phil. Stud.,* 1889, **5**, 601–617.

Smith, W. M. A methodological study of size-distance perception. *J. Psychol.,* 1953, **35**, 143–153.

Wohlwill, J. The development of "overconstancy" in space perception. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development,* Vol. I. New York: Academic Press, 1963. Pp. 265–312.

Zeigler, H. P., & Leibowitz, H. W. Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.,* 1957, **70**, 106–109.

(Received July 14, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 383–388

TRANSFER OF TRAINING AS A FUNCTION OF STIMULUS AND RESPONSE INTEGRATION OF PAIRED ASSOCIATES

M. J. HOMZIE, L. E. HARDACRE,[1] AND JON WEIMER[2]

University of Denver

A total of 180 Ss served in 1 of 2 paired-associate learning transfer studies in which the 1st task allowed Ss to integrate the stimulus and/or the response terms, form either backward or forward associations, or some combination of these factors which were appropriate for the subsequent transfer task. The experiments were designed to test the 2-phase conception of paired-associate learning as applied to transfer phenomena. This conception gave a good approximation of the 1st list and subsequent transfer task performance, and it was concluded that this notion can be fruitfully extended to the area of transfer.

Recently the two-phase conception of paired-associate learning has been used to integrate and explain various transfer of training phenomena (e.g., Ellis & Burnstein, 1960; Jung, 1963). Basically the two-phase conception of paired-associate learning is that *S* must learn to make the proper associative connection between the stimulus and response members as well as learn the response. For example, if the nonsense syllable KUZ were to be paired with CAT, then S would need to connect KUZ with CAT (the associative phase) as well as acquire KUZ as a correct response member (the response learning phase). It is assumed that these two processes are relatively independent, but that both could, and probably do, go on at the same time.

The two experiments to be reported were designed such that during the first task *S*s could have learned the associative connection between the stimulus and response members, or have integrated the response, or integrated the stimulus, or formed backward associations, or some combination of these factors which were appropriate for the subsequent transfer task. The two-phase conception of paired-associate learning gives rise to direct predictions as to the relative performance of such groups on the first task, and as to the effectiveness of the various first-task experimental conditions on the transfer task.

METHOD

Experiment I

Subjects.—The *S*s were 105 University of Denver undergraduates naive with respect to verbal-learning experiments who participated in the study as part of the introductory psychology course requirement. Five *S*s were eliminated (two Group SI; one from each of the other experimental conditions) for failure to meet the first task acquisition criterion. The remaining 100 *S*s were assigned to one of five conditions, the conditions differing from each other by the nature of the first list. All *S*s in every condition learned the same second list (List II) for the transfer task.

Verbal materials and experimental design.—Table 1 shows the verbal materials for the first task and the common transfer task for the five groups. The first column depicts the Stimulus-Integration (SI) group. Looking at the first verbal unit, I was paired with JH, and Y was paired with JH. The transfer task (List II) involved the learning of IY to JH. Therefore, in the transfer task *S*s in Group SI merely had to learn

---

[1] The second author was a National Aeronautics and Space Administration Trainee during the period of this investigation.

[2] The third author was a National Science Foundation undergraduate research participant during the period of this investigation under Grant GY-2068.

TABLE 1

VERBAL MATERIALS USED IN THE VARIOUS LISTS FOR EXP. I AND II

Initial Tasks						Transfer Task
Exp. I & II List SI	Exp. I & II List S & RI	Exp. I List RIB	Exp. I List S & RIB	Exp. I & II Control List	Exp. II List RI	Exp. I & II List II
I JH	I J	J IY	J I	5 #	IY J	IY JH
Y JH	Y H	H IY	H Y	11 ?	IY H	WX MK
W MK	W K	M XW	K W	7 $	XW K	AQ CF
X MK	X M	K XW	M X	9 &	XW M	PG TD
A CF	A C	C AQ	C A	4 %	AQ C	BN LZ
Q CF	Q F	F AQ	F Q	2 —	AQ F	SR VU
P TD	P T	T PG	T P	1 $	PG T	IY JH
G TD	G D	D PG	D G	8 #	PG D	WX MK
B LZ	N Z	L BN	Z N	3 —	BN L	AQ CF
N LZ	B L	Z BN	L B	10 %	BN Z	PG TD
S VU	R U	V SR	U R	12 ?	SR V	BN LZ
R VU	S V	U SR	V S	6 &	SR U	SR VU
(1)[a] 12	12	12	12	12	12	6
(2)[b] 2	1	2	1	1	1	2

[a] No. of associations to be learned.
[b] Size of response unit.

to integrate the stimulus unit (or use one element as the functional stimulus), as they already had the appropriate associative connections and the responses were already integrated. As can be seen in the next column, the Stimulus and Response Integration (S & RI) group had the appropriate associative connections from the first list, but in this case they had to integrate both the stimulus and response terms in order to master the transfer task. The next group, Response Integration Backward Association (RIB), initially acquired J to IY and H to IY. Glancing at List II it can be seen that the transfer task involves the integration of the response member and a backward association. The Stimulus and Response Integration Backward (S & RIB) group is identical to the S & RI group except that the associations formed are backward relative to the transfer task. Finally, a control group was run using stimulus and response items dissimilar from List II in order to control for warm-up and learning-to-learn effects.

The transfer task consisted of 12 bigrams which were paired to form six paired-associates. Each pair was presented twice in the list generating a 12-item list. The average frequency of occurrence in running text and the mean generated bigram frequency of the stimulus members and the response members were equated using the tables provided in Underwood and Schulz (1960).

Procedure.—The *S*s were given standard paired-associate anticipation method learning instructions. The verbal materials were presented on a Stowe memory drum at a 1.5 : 2 sec. rate. There were four different orders of each list to prevent serial learning of the responses by *S*s. Learning of the first list was carried to a criterion of 8 out of 12 correct. The experiment was terminated after 60 trials, if *S* had failed to meet the criterion. Upon *S* meeting the criterion for the first task, *S* then learned the transfer list until he had correctly anticipated the entire list (i.e., each of the six pairs twice).

The *S*s were assigned in blocks of 10 to the five groups. The first eight *S*s were randomly assigned two each to Groups SI, S&RI, RIB, and S&RIB. The ninth and tenth *S*s were placed into the control condition and received the same number of trials (i.e., yoked) on their first list as the average number of trials the other eight *S*s had spent in learning their respective initial tasks.

Experiment II

The anticipation method introduces the methodological difficulty of differential delay of knowledge of results for a Response Integration (RI) group as compared to the other experimental conditions. Thus, in order to run an RI group and also to extend the findings to the study-test trial (ST) procedure, Exp. I was partially replicated. In essence the procedures for Exp. I and II are identical except for the ST method and the inclusion of an RI group with forward associations. All cases where differences existed are noted in the following section.

Procedure and experimental design.—The second experiment involved four groups (SI, RI, S&RI, Control). Once again there were 20 *S*s in each group (2 of the 82 *S*s were eliminated for failure to meet the acquisition criterion on Task 1). The SI, S&RI, Control, and Transfer tasks were the same as in Exp. I (see Table 1), except that *S*s learned by the ST procedure, i.e., alternating study trials with test trials. The verbal units were presented for 1.5 sec. on study trials, and *S*s were allowed 1.5 sec. to respond on the test trials. Turning to the RI group (see Table 1, sixth column), *S*s were presented (as an example) with IY-J and IY-H as their first task. The transfer task consisted of acquiring IY-JH which involves just response integration (assuming that *S*s had completely mastered the initial learning situation).

The experiment was terminated after 30 test trials if *S* had failed to meet the acquisition criterion of 8/12 correct on the first list. In addition, *S*s in Group RI had to give six *different* correct responses. This was to insure that *S*s were learning both elements of the response rather than learning a single response and giving it twice during the test trials. Surprisingly, *S*s in this group had little, if any, difficulty in understanding the instructions and in completing their initial task.

RESULTS AND DISCUSSION

First list.—The results in terms of the trials to criterion (TTC) and errors (E) data for the initial task are summarized in Table 2. The E analysis which includes both overt errors and failures to respond appears in parentheses. Simple one-way analyses of variance were run for both Exp. I and II on the two dependent measures. Since the variances appeared to be relatively homogeneous the within-groups mean square was used in applying the multiple *t* tests. Table 2 shows the results of this analysis.

An attempt will be made to explain the results *solely* in terms of the two-phase conception of learning paired-associates. Note that (see Table 1) all the groups had the same 12 letters as the response members on the first task, but the groups differed in the *size* of the response unit that had to be integrated. It is assumed that learning the same response to two different stimuli constitutes two separate associations, consequently all the groups had an equal number of associations

TABLE 2

SUMMARY OF THE TTC AND ERROR ANALYSIS FOR THE FIRST LIST FOR EXP. I AND II

	SI TTC (E)	RIB TTC (E)	S & RI TTC (E)	S & RIB TTC (E)
Exp. I[a]				
$\bar{X}$s	23.8 (209.9)	20.8 (173.5)	16.8 (133.4)	14.3 (119.3)
*SD*s	8.6 (81.1)	8.2 (66.5)	9.6 (79.2)	13.9 (84.9)
SI		(<.05)[c]	<.01	<.01
RIB			<.05	<.01
S & RI				
	RI TTC (E)	S & RI TTC (E)	SI TTC (E)	
Exp. II[b]				
$\bar{X}$s	5.6 (36.1)	6.7 (45.7)	9.1 (68.5)	
*SD*s	3.4 (27.0)	5.1 (39.8)	3.8 (32.5)	
RI			<.01	
S & RI			<.05 (<.01)[c]	

[a] $F\ (3, 76) = 4.21,\ p < .01$; $F\ (3, 76) = 5.21,\ p < .01$.

[b] $F\ (2, 57) = 3.45,\ p < .05$; $F\ (2, 57) = 4.70,\ p < .05$.

[c] The TTC and error analysis resulted in a different significance level for the indicated comparison.

TABLE 3

SUMMARY OF THE TTC AND ERROR ANALYSIS FOR THE TRANSFER TASK FOR EXP. I AND II

	SI TTC (E)	RIB TTC (E)	S & RIB TTC (E)	S & RI TTC (E)	Control TTC (E)
Exp. I[a]					
$\bar{X}$s	6.4 (17.4)	11.7 (59.6)	14.2 (82.6)	17.7 (112.2)	14.9 (77.1)
*SD*s	4.7 (17.7)	5.8 (36.4)	6.6 (40.2)	14.4 (107.7)	4.1 (31.3)
SI		<.01	<.01	<.01	<.01
RIB			(<.05 one-tailed)[c]	<.01 (<.05)[c]	
S & RIB					
S & RI					

	SI TTC (E)	RI TTC (E)	S & RI TTC (E)	Control TTC (E)
Exp. II[b]				
$\bar{X}$s	3.2 (4.9)	7.2 (30.2)	8.2 (42.8)	7.3 (28.7)
*SD*s	1.5 (4.3)	4.8 (25.6)	5.4 (36.0)	6.2 (32.0)
SI		<.01	<.01	<.01
RI				
S & RI				

[a] $F(4, 95) = 5.27, p < .01; F(4, 95) = 7.30, p < .01.$

[b] $F(3, 76) = 4.02, p < .01; F(3, 76) = 6.32, p < .01.$

[c] The TTC and error analysis resulted in a different significance level for the indicated comparison.

(i.e., 12) to learn on the first task. Thus, turning to Exp. I, Groups S&RI and S&RIB should be equivalent to each other (size of response unit = 1) and superior to Groups SI and RIB which also should be equal to each other (size of response unit = 2). These results were obtained but for one exception; Groups SI and RIB were statistically different from each other for the error analysis.

A similar interpretation of Exp. II would lead to the prediction that Groups RI and S&RI should be equivalent in performance (size of response unit = 1) and both of these groups should be superior to Group SI (size of response unit = 2). As can be seen in Table 2 the results support the two-phase model.

Transfer task.—Table 3 summarizes the transfer task data. The analysis is similar to that of the first list: simple one-way analyses of variance and subsequent multiple *t* tests among the groups. In that the variances appeared to be heterogeneous, the error terms were calculated directly in applying the *t* tests.

Three simplifying assumptions were made in order to straightforwardly apply the two-phase conception to the transfer data. (*a*) In that the stimulus member is always presented on the memory drum, response integration will be a more potent factor and will lead to poorer performance than stimulus integration. (*b*) A prior backward association will be just as effective as a forward association (i.e., associative symmetry, cf. Asch & Ebenholtz, 1962; Horowitz, Norman, & Day, 1966). (*c*) All three factors (stimulus integration, response integration, and associative connection) will combine to determine the second-list performance. Note that the first and third statements are direct implications of the two-phase model.

From the above it can be predicted for Exp. I: SI > RIB > S & RI = S & RIB (see Assumptions *a*, and *b*, *c*, and *b*, respectively). Table 3 reveals that the only comparison which is contrary to the above predictions is that the performance of Group RIB was not statistically superior to that of Group S & RIB (the differences were in the appropriate direction and reached significance for the E analysis if directionality of the prediction is considered). Turning to Exp. II, the predictions are: SI > RI > S&RI (see Assumptions *a* and *c*, respectively). As Table 3 indicates, the only exception to the above predictions is that Group RI was not statistically superior in performance to Group S&RI (although the results were in the appropriate direction).

The two-phase conception of paired-associate learning also has implications as to the relative magnitude of the differences between groups. The difference in performance between an SI group and an RI group is that the latter group needs to learn to integrate the response, whereas, the former group is only required to integrate the stimulus (which is a minor problem in that the stimulus is available on the memory drum and in addition *S* need use only one element of the stimulus term as the functional stimulus). The difference in performance between an RI group and an S&RI group is that the latter group has the added task of integrating the stimulus terms. Thus the difference between an SI group and an RI group is primarily that of response integration, while the difference between an RI group and an S&RI group is that of stimulus integration. Recall that response integration is assumed to be a more potent factor than stimulus integration (Assumption *a*), thus the implications for Exp. II are: SI – RI > RI – S&RI, and for Exp. I: SI – RIB > RIB – S&RIB; SI–RIB > RIB – S&RI. Table 3 indicates that the only differences which were not in the appropriate direction are those for the last prediction for Exp. I (i.e., the S&RI group's performance was too low). A post hoc inspection of this group's data indicates that the performance of one *S* deviated extremely. This *S* committed 501 errors and took 64 trials to master the transfer task (greatly contributing to the unusually large standard deviation of this group). Eliminating this *S* from the analysis reduces the S&RI group's means to 15.2 and 91.7 for the TTC and E data, respectively, which would bring the data for this group in line with the prediction.

Thus, for the initial and the transfer tasks in Exp. I and II, of the 36 possible comparisons among the experimental groups for the two dependent measures, only five comparisons were not in agreement with the model and four of these were in the appropriate direction. In addition two of the three predictions as to the relative differences between groups were in the appropriate direction for both dependent measures, and the one failure could possibly be attributed to sampling error.

Degree of first-list learning effects.—The derivations from the two-phase conception of paired-associate learning tend towards predictions that the groups which are expected to perform the worst on the first list will be the best on the transfer task (e.g., Group SI was predicted to be one of the poorest performers for the initial task and the best performing group on the second list). It would be expected that the transfer-task performance of the groups which have had the greatest number of trials on the first list would be somewhat superior than the other groups due to the effects of overlearning of the easy items, learning-to-learn, and warm-up. In that the theorizing for the first list was accurate, these factors tend to produce the predicted results for the subsequent transfer task, but it seems doubtful that the positive findings of this study can be explained solely by the effects of degree of first-list learning.

Relative transfer effects.—It should be noted that in Exp. I and II, the S&RI groups did not perform better than the Control groups indicating that there was interference transferred from the initial list. Undoubtedly, the first task structured the transfer task in such a way as to interfere with the idiosyncratic learning habits of some of the *S*s. Turning to a specific example of interference from the first list, 30.7% of the RI group's errors were classified as "wrong integration" (the response term's elements were reversed, e.g., HJ instead of JH). In fact, 10.8% of *all* the responses this group made were wrong integration errors (WIE), whereas, for the Control group only .7% of its total responses were WIE, which constituted 2.2% of its total errors. In a similar vein, 21.3% of the errors committed by Group RI were mixed errors (one of the response elements was appropriate to a different stimulus term, e.g., JK instead of JH). Only 12.3% of the Control group's errors were of this type.

Conclusion.—Returning to the two-phase conception of paired-associate learning, this notion gave a good approximation of the first list and subsequent transfer task performance. The results of this investigation and of other studies (e.g., Ellis & Burnstein, 1960; Jung, 1963) suggest that the two-phase conception of paired-associate learning can be fruitfully extended to transfer phenomena.

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principle of associative symmetry. *Proc. Amer. Phil. Soc.*, 1962, **106**, 135–163.

ELLIS, H. C., & BURNSTEIN, D. D. The effect of stimulus similarity and temporal factors in perceptual transfer of training. Technical Report No. 1, 1960, Sandia Corporation, Albuquerque, New Mexico.

HOROWITZ, L. M., NORMAN, S. A., & DAY, R. S. Availability and associative symmetry. *Psychol. Rev.*, 1966, **73**, 1–15.

JUNG, J. Effects of response meaningfulness (m) on transfer under two different paradigms. *J. exp. Psychol.*, 1963, **65**, 377–384.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received July 15, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 389–392

HUMAN SALIVARY CONDITIONING: EFFECT OF UNCONDITIONED-STIMULUS INTENSITY[1]

BEN W. FEATHER, FREDERICK C. DELSE, AND MARION R. BRYSON

Duke University

4 groups of 14 high-school students each were administered 8 classical salivary conditioning trials. Each group received 1 of 4 concentrations of acetic acid employed as the unconditioned stimulus: .03%, .3%, .5%, and 1.5%. 2 groups for which the CS and US were unpaired received .3% and 1.5% acetic acid. Amount of salivation during the 10-sec. CS-US interval was compared with amount of salivation during comparable dummy periods. Significant levels of conditioning were demonstrated at all 4 US concentrations. There was a significant ($p = .04$) linear relationship between US intensity and magnitude of conditioned response. US intensity, however, did not affect significantly the probability of response. Pseudoconditioning was not demonstrated by the unpaired groups.

There are only a few English language reports of human salivary conditioning and of these, fewer still are based on experiments which utilized adequate methods and controls in sufficiently large *S* samples. Optimal values are unknown for any of those parameters which are considered relevant to other classical conditioning. The present study concerns one such parameter, the effect of unconditioned stimulus (US) intensity on anticipatory conditioned salivation.

Warstler and Ost (1965) investigated the effect of three US intensities on salivary conditioning in the dog. They concluded that, within limits, the magnitude of salivary conditioned responses (CR) varied directly with intensity of an acid US, but that this relationship was inverted at high intensities. Otherwise, their findings were comparable to those cited by Razran (1957) and Gantt (1938) for salivary conditioning in dogs. Warstler and Ost concluded that the inversion at the highest US intensity might have arisen from differences in prestimulus salivary rates.

A major problem encountered in studying salivary conditioning results from the fact that salivation occurs more or less continuously and at varying rates. A salivary CR, therefore, most often is defined as some rate of salivation to a CS compared to the rate in some other time segment. Customarily, the time segment selected to represent a base line against which to compare conditioned salivation has been an interval immediately preceding the CS at some fixed time after the preceding trial. This practice has the disadvantage of combining markedly different (inter- and intra-*S*) prestimulus rates. This problem becomes especially serious when the variable investigated produces systematic effects on prestimulus rate. Since US intensity certainly affects prestimulus salivary rate when the usual range of intertrial intervals (ITI) is employed, it was decided to administer conditioning trials contingent upon a specified salivary rate, rather than at predetermined time

---

[1] This investigation was supported by grants from the Duke Endowment; the Research Foundation of the National Association for Mental Health, Incorporated; and by three grants from the National Institute of Mental Health: a Public Health Service Career Development Award, K3-MH-19, 523; MH 11071; and MH 11549.

intervals. While this strategy should reduce the amount of spurious variability in the magnitude of CRs, it obviously raises some questions of controls. These questions will be discussed in a subsequent section.

METHOD

Subjects.—The *S*s were 42 male and 42 female high-school students ranging in age from 14 to 19. They were recruited by advertising and paid for their services.

Apparatus.—The saliva collecting, measuring, and recording apparatus has been described in more detail elsewhere (Feather & Wells, 1966). Saliva is collected by a double-chambered stainless-steel capsule placed over Stensen's duct in the left cheek. The outer chamber is connected to a source of vacuum and holds the capsule in place. The inner chamber rests directly over Stensen's duct and is connected via semirigid polyethylene tubing to a flask containing carbon tetrachloride. Sufficient vacuum is applied to this flask to provide approximately atmospheric pressure at the salivary duct. Saliva at the exit of the duct displaces an equal volume of carbon tetrachloride through a dropper needle. Drops of 1/480cc volume interrupt a light beam aimed at a photoelectric cell. The interruption of the light beam by the drops provides a pulse which is recorded by a Grass Model 7 polygraph. In addition, a second-by-second drop count is made by electromechanical counters during certain specified intervals.

The US, acetic acid, was administered to the left lateral margin of the tongue via polyethylene tubing attached to the outside of the parotid capsule. The quantity and duration of the US was controlled by the use of solenoid valves and electronic timers.

Swallowing, tongue movements, and motor activity of the mouth were monitored by electromyographic recording. Two recording electrodes were placed 1 cm. on either side of the midline of the submental region and a reference electrode was placed on the left mastoid area. Further control over motor activity was achieved by instructing *S*s to swallow only when a sign in front of them reading "YOU MAY SWALLOW NOW" was lighted. The CS was a 600-cps tone at approximately 60 db. (re .0002 dyne/cm^2) interrupted once per sec. and administered by an Eico code practice oscillator via Monarch earphones. The *S*s were isolated from *E* and apparatus in a separate, sound attenuated chamber.

Procedure.—The *S*s were assigned to four conditioning and two unpaired-control groups, each containing 14 *S*s, 7 male and 7 female. USs for the four conditioning groups were 3 cc of acetic acid of the following concentrations and acidities: .03% (pH 3.6), .3% (pH 3.3), .5% (pH 3.1), and 1.5% (pH 3.0). These acetic acid solutions were selected at these pHs to provide approximately equal increments of hydrogen ion concentration. The *S*s in the two unpaired-control groups received USs of 3 cc of .3% and 1.5% concentration, respectively. Chauncey and Shannon (1959) demonstrated that the magnitude of the parotid secretory response to acid is a linear function of the hydrogen ion concentration. The lowest concentration of acetic acid was only slightly above sourness threshold for most *S*s and the highest concentration was judged definitely noxious by most *S*s. The sequence and duration of all stimuli were controlled by automatic programming equipment.

The *S*s were instructed to fast for the 6 hr. preceding experimental sessions. At the beginning of a session *S* was seated in the experimental chamber and the apparatus was briefly explained while it was being attached. Prerecorded instructions requested *S* to restrict swallowing and mouth and tongue movements to times when the "YOU MAY SWALLOW NOW" sign was lighted. This sign was lighted for both conditioning and unpaired-control groups at the offset of the US and remained lighted for 30 sec. Motor activity and salivation were continuously monitored throughout experimental sessions.

Each *S* in the conditioning groups received eight trials. A trial consisted of a 13-sec. interrupted tone overlapped by the administration of 3 cc of acid in the last 3 sec., thus providing an interstimulus interval of 10 sec. These trials were administered by appropriate timing and programming equipment at an ITI directly dependent upon ongoing salivary rate. No trial was given for 90 sec. following the preceding trial. After 90 sec. a trial was given automatically on the first occasion on which the time between two drops of saliva was at least 3 sec. Thus all conditioning trials occurred at comparable prestimulus salivary rates, while the ITI varied from trial to trial and from *S* to *S*.[2]

---

[2] Controlling the prestimulus salivary rate in this manner introduces a systematic

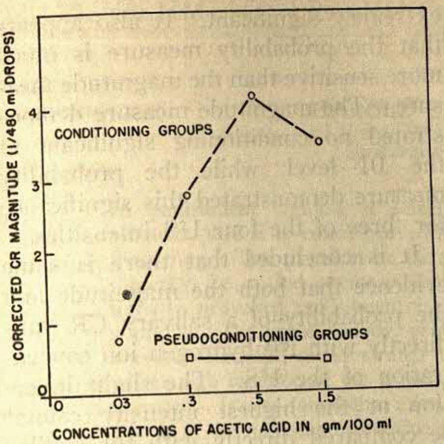


Fig. 1. Mean CR magnitude for each concentration and mean magnitude of unpaired control responses.

To establish an expected value of salivary rate against which CR periods could be compared, similar periods in which salivation could be measured were interspersed following every second trial. During these 10-sec. periods *S* received no stimulation but his salivary rate was measured automatically, and recorded both on the polygraph record and on electromechanical counters. These dummy trials were given by the same criteria as were conditioning trials.

Each unpaired-control *S* received eight 13 sec. CSs and eight 3 sec. USs, unpaired. The presentation of these CSs and comparable dummy measuring periods was accomplished in the same way as for conditioning *S*s.

Results and Discussion

TABLE 1

Analysis of Probability Measure by the Normal Approximation to the Binomial

Concentration	Proportion of Positive Differences	t	p (2-tailed)
.03	.64	1.80	.08
.3	.71	3.00	.003
.5	.73	3.13	.002
1.5	.84	4.50	.0001
Control .3	.52	.17	.90
Control 1.5	.64	1.80	.08

The major question to be answered in this experiment was the following: "Does the 'amount' of salivary conditioning which occurs in the experiment depend upon the concentration of acid in the US?" "Amount" may be defined in two ways. First, one can measure it by the quantity of drops by which the experimental trials exceeded the dummy trials. One can also measure "amount" of conditioning by the fraction of the experimental trials in which the number of drops exceeded the number of drops in the corresponding dummy trial. The first of these will be called the magnitude measure, the second, the probability measure.

Analysis of data by magnitude measure.—A t test was applied to determine whether conditioning did indeed take place when the magnitude measure was used. There was a significantly higher salivary rate in the experimental trials than in the dummy trials for all four conditioning groups, $p < .05$.[3] In the unpaired controls' trials, this difference was not significant at either concentration, $p = .40$.

Figure 1 shows the mean CR magnitude for each concentration and the mean magnitude of unpaired control responses (called pseudoconditioning groups in the figure).

Analysis of variance failed to show a significant difference between the amount of conditioning at the four concentrations, $F = 1$. However, a trend analysis showed that there was a sig-

---

lengthening of ITI, for the higher US intensity groups. Although there was a great deal of variation in ITI within each conditioning group, the average ITI was approximately 30 sec. longer for the 1.5% group than for the .03% group.

[3] There were two extreme values at the .3% concentration. Application of a test for outliers (Natrella, 1963) indicated that they could be treated as outliers, $p < .01$. Thus, they were omitted from this and succeeding magnitude measure analyses.

nificant linear trend, $F\ (1, 50) = 4.72$, $p < .05$.

Omitting the conditioning trials of the .03% and the .5% groups and including the unpaired control trials of the .3% and the 1.5% groups, a two-way analysis of variance was performed. The experimental vs. the unpaired control variable produced a significant value, $F\ (1, 50) = 4.55$, $p < .05$. There was not a significant difference between the two US intensities, $F = 1$, nor was the interaction significant, $F\ (1, 50) = 1.14$.

Analysis of data on probability measure.—Of all conditioning trials, only those for which the salivary rate in the experimental trial differed from the rate in the corresponding dummy trial were used in this analysis. Also, since there were two conditioning trials for each dummy trial, the salivary rate for the conditioning trial immediately preceding and the one immediately following a dummy trial were averaged to obtain a single rate for comparison with the dummy-trial rate.

The normal approximation to the binomial distribution was used here to determine whether there were significantly more positive (experimental exceeded dummy) differences than there were negative differences. The results of the analysis are shown in Table 1.

Although, using the studentized-range test (Winer, 1962, p. 77) the difference between the largest and smallest proportion was not quite significant; $p = .06$, examination of the progressively larger proportions in Table 1 lends some support to the conclusion that US intensity has an effect on salivary conditioning. At the higher US intensities the conditioning was extremely significant. It also appears that the probability measure is much more sensitive than the magnitude measure. The magnitude measure demonstrated no conditioning significant at the .01 level, while the probability measure demonstrated this significance for three of the four US intensities.

It is concluded that there is some evidence that both the magnitude and the probability of a salivary CR vary directly with the hydrogen ion concentration of the US. The slight inversion at the highest intensity cannot be compared directly with the similar finding by Warstler and Ost (1965) since the acid concentrations were not the same. Pseudoconditioning was not demonstrated.

REFERENCES

Chauncey, H. H., & Shannon, I. L. Parotid gland secretion rate as a method for measuring response to gustatory and masticatory stimuli in humans. *Rep. Sch. Aviation Med.*, Randolph AFB, Texas, 1959.

Feather, B. W., & Wells, D. T. Effects of concurrent motor activity on the unconditioned salivary reflex. *Psychophysiology*, 1966, **2**, 338–343.

Gantt, W. H. The nervous secretion of saliva: The relation of the conditioned reflex to the intensity of the unconditioned stimulus. *Amer. J. Physiol.*, 1938, **123**, 74–75.

Natrella, M. G. *Experimental statistics. Nat'l bur. of standards handbook 91*, 1963.

Razran, G. The dominance-contiguity theory of acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.

Warstler, H. E., & Ost, J. W. P. Classical salivary conditioning in the dog: effects of three US intensities. *J. comp. physiol. Psychol.*, 1965, **60**, 256–259.

Winer, B. J. *Statistical principles of experimental design.* New York: McGraw-Hill, 1962.

(Received July 15, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 393–399

PUNISHMENT TRAINING WITH AND WITHOUT AN ESCAPE CONTINGENCY

HAROLD LEITENBERG[1]

University of Vermont

In punishment training a selected response produces an aversive stimulus, and E is concerned with the subsequent decline in response frequency. The punishment paradigm may contain an escape contingency, i.e., a 2nd response distinguishable in kind from the 1 that initially provoked punishment is required to terminate the punishing stimulus. The present study, using rats as Ss and bright light as an aversive stimulus, demonstrated that the effectiveness of punishment varies as a function of the presence or absence of such an escape contingency. It was found that: (a) punishment had greater suppressive effects in the no-escape condition than in the escape condition; (b) response frequency was more resistant to recovery following removal of punishment in the no-escape condition than in the escape condition.

Theorists have hypothesized that the suppressive effects of punishment simply reflect the acquisition of avoidance responses which compete with the original response being punished (e.g., Dinsmoor, 1954, 1955; Mowrer, 1960). Supposedly the punished response occurs less frequently because competing avoidance response(s) are established in its place. As Solomon (1964) points out, this mirror-image view of avoidance conditioning and punishment training suggests that "new experimental relationships discovered in the one experimental setting should tell us how to predict a new empirical event in the other setting." For instance, variables that cause rapid acquisition of an avoidance response should cause rapid response suppression in the punishment paradigm, and vice versa.

Often in avoidance conditioning the same response, whether it be pressing a bar or jumping from one chamber of a shuttle box to another, serves both an avoidance and an escape function. When made in time, the bar press will prevent shock (avoidance); if the shock comes on, however, the bar press will turn the shock off (escape).

When the escape response and the avoidance response are the same, such an escape contingency slightly facilitates avoidance learning (Marx & Hellwig, 1964; Mowrer & Lamoreaux, 1946). It sometimes even helps to maintain avoidance responding (Hearst & Whalen, 1963), although this effect is not always seen (Malott, Sidley, & Schoenfeld, 1963). On the other hand, when the escape response and the avoidance response are topographically different, the presence of an escape contingency significantly retards avoidance learning (Mowrer & Lamoreaux, 1946).

The role of an escape contingency in the punishment situation has so far attracted very little experimental analysis (Leitenberg, 1965). In the punishment paradigm such an escape con-

---

[1] These experiments were carried out at Indiana University while the author had a predoctoral fellowship (1-F1-MH-23, 282-01) from the National Institutes of Health. Preparation of this manuscript was supported by a National Science Foundation institutional grant (GU-1505) to the University of Vermont. Reprints can be obtained from the author, Department of Psychology, University of Vermont, Burlington, Vermont.

tingency is present if an instrumental response distinguishable in kind from the one that initially produced the aversive stimulus is required to terminate the aversive stimulus. The escape contingency comes into effect after the punishing stimulus has been produced by the "incorrect" response, and it does not alter the conditions responsible for onset of punishment (cf. Azrin, Hake, Holz, & Hutchinson, 1965). It does, however, permit *S* self-control of each punishment duration; the sooner the escape response is made, the sooner the punishing stimulus is terminated.

This experiment is designed to determine if an escape contingency has the same effect in the punishment paradigm as it has in the avoidance paradigm. It can be expected that an escape contingency which decreases avoidance responding (Mowrer & Lamoreaux, 1946), will also decrease the suppressive effects of punishment.

EXPERIMENT I

Method

Subjects.—The *S*s were 60 male albino rats, approximately 90–120 days old at the start of the experiment. Food and water were provided on an ad lib basis in their home cages.

Apparatus.—The apparatus is fully described in Leitenberg (1965). Onset of a clear 300-w. bulb located 18 in. above the floor of the box was the aversive stimulus. The amount of illumination as measured by a Weston photronic ftc. meter at the level of the rat's head in the experimental box was 200 ftc. Because the light was centered, it did not cast any shadows on the floor of the box. The floor and side walls were painted silver. An exhaust blower, heat absorbent glass, and the usual short time the light was on prevented the heat generated by the bulb from reaching the rat. Over a session, the temperature in the box never rose more than 2° above the temperature in the room. At one end of the sound insulated box was a 10½-in. wide, 4-in. long silver painted bakelite platform, raised ½ in. above the surface of the box floor. At the opposite end of the box was a 5-in. wide, 3-in. long aluminum lever located 2½ in. above the floor. The two manipulanda were 7 in. apart, and the width of the box was 11½ in. Depression of the platform was the response punished by onset of the bright light, and depression of the aluminum lever was the escape response. Thirty-gram pressure was required to operate the switch associated with the platform, while 20-gm. pressure was required to operate the switch associated with the lever. A tape programmer controlled onset of the light for the uncontingent light group (see below), and electromechanical timers controlled light duration for the no-escape condition. Electromechanical counters and running time meters in connection with appropriate relay circuitry were used to record response frequency, response duration, and response latency.

Design and procedure.—The 60 *S*s were assigned at random, 15 to each of the four groups designated below. For Group 1, onset of light followed depression of the bakelite platform, and termination of light was contingent on depression of the aluminum lever. The duration of punishment varied as a direct function of the latency of the escape response.

For Group 2, light onset was again contingent on the platform response, but light duration was fixed and determined by *S*s in Group 1. Although the aluminum lever was present in the box while *S*s in Group 2 were being run, depression of this manipulandum had no programmed escape consequence. The mean-light duration per light onset was computed for each *S* in Group 1, for each session, and these means determined the time the light stayed on for *S*s in Group 2. For instance, if the mean light time for *S* 1 in Group 1 in Session 1 was 20 sec., then every platform response made by *S* 1 in Group 2 in Session 1 in the absence of light was followed by 20 sec. of light. Since it was possible for the platform to be pressed more than once by *S*s in Group 1 before the escape response was made and the light was turned off, there was a choice between equating the yoked *S*s on mean light duration per light-producing platform response or equating them on mean light duration per every platform response. The former procedure was considered more desirable.

For Group 3, neither depression of the platform nor depression of the lever had any programmed consequence. The *S*s in this group never experienced onset of the bright light. The purpose of this condition was to determine if the frequency of the nonreinforced operants used in the present experiment would change over sessions without punishment or escape contingencies.

For Group 4, onset of the light was not contingent on the platform response, but termination of the light was contingent on depression of the aluminum lever. The number and temporal sequence of light onsets was yoked to that produced by *S*s in Group 1. Group 4 served two control functions simultaneously: (*a*) The effects of response independent light on platform pressing frequency could be assessed. (*b*) When the escape contingency was in effect, the time required to make the escape response, although insubstantial, might cause less time to be available to make the platform response. If this were true, then Group 4 should press the platform less frequently than Group 3, even though the platform response was not punished for either of these groups.

There were nine sessions, each lasting 15 min. The number nine was selected because earlier work (Leitenberg, 1965) had already demonstrated that asymptote performance is reached in about this time. Animals were run on alternate days and the punishment and escape contingencies for the various groups were introduced in the first session.

Results

Median platform responses are plotted as a function of sessions in Fig. 1. Although platform-response frequency declined somewhat over sessions for all groups, the decline was considerably greater when punishment followed this response (Groups 1 and 2) than when punishment was either absent (Group 3) or not response contingent (Group 4). An analysis of variance performed on median response frequency yielded a significant F ratio, $F\ (3, 56) = 15.24$, $p < .01$. A Duncan multiple-range test indicated that Group 1, trained under both punishment and escape conditions, made significantly more responses than Group 2, trained under conditions of punishment without escape ($K = 2$, $r_2 = 4.92$, $p < .05$). Groups 3 and 4 did not significantly differ from each other, but each made significantly more platform responses than either Group 1 or Group 2 ($p < .05$).

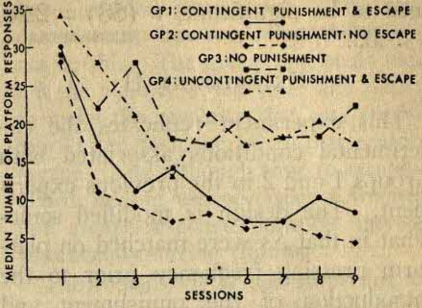


Fig. 1. Median number of platform responses in successive sessions; Exp. I.

In the usual escape-training situation, the number of escape responses increases as training progresses. In the present study, however, there was no such increase probably because as training progressed, the number of punishments to be escaped declined. Another contributing factor may have been the unusually high operant level of the lever response.

On the other hand, there are several indications that *S*s who could escape were indeed doing so. The time *S*s took to press the aluminum lever following a platform response was recorded. Since depression of the lever had no programmed escape consequence for Groups 2 and 3, and since for Group 4 there was nothing to escape following platform depression, this time should be shortest for Group 1. The median time in seconds for Groups 1 was 6 and the median time in seconds for Groups 2, 3, and 4 combined was 20. An orthogonal F test performed on this median data yielded a significant F ratio, $F\ (1, 56) = 5.25$, $p < .05$. Also, the mean of the median number of lever responses for the two escape groups was greater than for the two no-escape groups (27.17 vs. 20.37). A t test showed this difference to be significant, $t\ (58) = 2.28$, $p < .05$.

EXPERIMENT II

This experiment replicates the experimental conditions associated with Groups 1 and 2 in the previous experiment. The design is modified somewhat in that *S*s were matched on platform pressing frequency prior to the introduction of the punishment and escape contingencies. In addition, the present experiment was designed to study recovery from the effects of punishment as a function of prior escape and no-escape conditions.

Method

Subjects.—Thirty male albino rats approximately 90–120 days old were initially tested, but because of difficulty in matching, only 24 rats were run in the experiment proper. One pair of rats had to be discarded during the 8-day rest period because one animal died.

Apparatus.—The apparatus was the same as in Exp. I, except that the aluminum lever was removed from the box when *S*s in the no-escape group were run. This was to prevent any adventitious reinforcement of the lever response for this group.

Procedure.—The first session was used to measure operant levels of the platform response in the absence of punishment. These values were then used to form matched pairs of *S*s. Within each pair, *S*s were randomly assigned to either the punishment with escape or punishment without escape conditions.

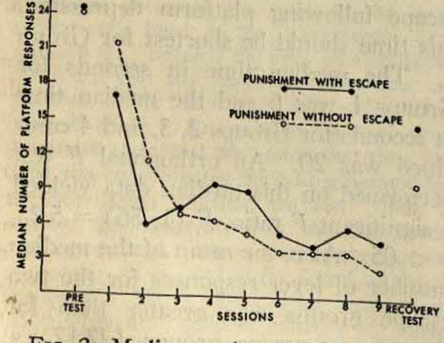


FIG. 2 Median number of platform responses in succesive sessions; Exp. II.

As in Exp. I, a yoking procedure was employed to equate mean duration of light per light-producing platform response within each pair of *S*s. The punishment contingency was in effect for a total of nine sessions. Each session lasted 15 min., and *S*s were run on alternate days. Following the last training session, *S*s were kept in their home cages for 8 days and on the ninth day retested in the experimental box without punishment.

Results

The major finding of Exp. I was replicated in the present experiment. Response frequency was suppressed to a greater degree by punishment without escape than by punishment with escape (see Fig. 2). A t test for matched pairs performed on median number of platform responses made during the nine training sessions indicated that this difference was significant, $t\ (11) = 2.66$, $p < .05$. This difference was also maintained over the 8-day rest interval. In the recovery test session the no-escape group made significantly less responses than the escape group, $t\ (10) = 3.83$, $p < .01$.

Following the removal of punishment, platform responding partially recovered; for both groups a sign test showed that response frequency during the recovery session was significantly higher than during the last training session ($p < .01$). Undoubtedly, both the rest interval and the fact that punishment was absent during the recovery session account for response recovery. What is more interesting, however, is the suggestion in Fig. 2 that the degree of recovery was less for the no-escape group than for the escape group. A t test for matched pairs was performed on difference scores obtained by substracting platform-response frequency in the last training session from platform-response frequency in the recovery-test session.

The degree of recovery was significantly less for the punishment without escape group than for the punishment with escape group, $t\ (10) = 2.67,\ p < .05$.

DISCUSSION

The results of Exp. I and II support previous findings regarding the aversive properties of bright light with rats. For instance, Hefferline (1950) previously demonstrated avoidance of bright light; Keller (1941) demonstrated escape from bright light; and Leitenberg (1965) demonstrated punishment with bright light.

This study, however, was addressed to a more specific question: namely, does the suppressive effect of punishment vary as a function of the presence or absence of an escape contingency?

The results indicate that punishment of a response is most effective when escape is prevented. This corresponds with the effect of an escape contingency in the avoidance situation; and it confirms the notion that results in the punishment paradigm can be predicted on the basis of results in the avoidance situation. Mowrer and Lamoreaux (1946) found that avoidance responding was increased when escape was prevented—so long as the avoidance response and escape response were different. In the present study, therefore, it was expected that punishment would be more effective when escape was prevented—so long as the hypothetical competing avoidance response(s) and the escape response were different. It may be questioned whether there was indeed a topographical and temporal difference between these two responses in this experiment.

This question cannot be answered definitively since in the punishment situation avoidance responses are unmeasured theoretical constructs. The escape response in the present experiment could have had an avoidance function. The *S*s who were able to escape might have spent all their time around the escape manipulandum. Thus, they would have been unable to press the platform and would have avoided punishment. The results demonstrate, however, that the rats did not do this. The group that could escape left the vicinity of the lever and returned to make the response that was punished more often than the group that could not escape. And although time to escape declined, neither escape-response frequency nor escape-response duration increased significantly during training.

Theoretically it is not surprising that the escape response in this experiment had no simultaneous avoidance function. The reinforcement supposedly responsible for the development of avoidance behavior in the punishment situation stems from the removal of proprioceptive cues associated with making the punished response (Dinsmoor, 1954). These cues are said to acquire negative reinforcing power because they immediately precede and set the occasion for the onset of the primary aversive stimulus. Such proprioceptive cues, however, were not available for removal in the present experiment until *S* approached the platform and began the chain of responses specifically associated with platform depression. A potential avoidance response must break this chain in order to be reinforced, a criterion which the escape response of the present experiment cannot meet without stretching the chain concept beyond reasonable limits. In this regard it seems worth noting that as training progressed, the rats were often observed to approach the platform, raise their front paws above it, and then draw back without stepping on the platform. Unfortunately, no systematic attempt was made to record this activity.

The avoidance-response theory of punishment implies that recovery from the suppressive effects of punishment is due to extinction of conditioned avoidance responses, and that variables which retard such extinction will also retard recovery. Therefore, given the findings of Marx and Hellwig (1964) that extinction of an observable avoidance response is slower following inescapable shock, it follows that recovery from punishment will also be slower after training with an inescapable aversive stimulus. This hypothesis receives support from the data in Exp. II: The formerly punished response was significantly more resistant to recovery in the no-escape group than in the escape group. (Although it seems unlikely, there is a possibility that this observed difference in the amount of increase in response frequency during the recovery phase does not reflect the prior presence or absence of the escape contingency per se; instead, it may reflect the fact that the escape and no-escape groups enter the recovery period with different rates of responding. That is, no matter what variable causes a higher response frequency at the end of punishment, the higher it is, within limits, the greater will be the subsequent amount of response increase during recovery. Presumably this would be a nonlinear relationship since in many instances where response rate is high at the end of punishment there would be little room left for further recovery when punishment is removed.)

This recovery from punishment data can be interpreted in terms of Marx and Hellwig's (1964) data on extinction of avoidance responding, because the recovery-test session was actually an extinction session. The punishing stimulus, the presumed primary reinforcer for the hypothetical competing avoidance response(s), had been removed. The escape contingency would be expected to have the same effects if test sessions had been run without an intervening rest period.

Marx and Hellwig speculated (*a*) that the aversive stimulus has greater motivating (fear?) properties (as demonstrated by defecation, etc.) when escape is prevented; and (*b*) that this greater motivation accounts for the greater resistance to extinction in the avoidance without escape condition. Similarly, it may account for the greater resistance to recovery and more complete suppression in the punishment without escape condition of the present experiments.

There are two independent sources of support for this motivation hypothesis. In a CER study, Mowrer and Viek (1948) found that a CS preceding inescapable shock caused greater suppression of eating than a CS preceding escapable shock. Mowrer and Viek conclude that a greater degree of fear is conditioned to stimuli preceding shock whose duration cannot be self-controlled than to stimuli preceding shock whose duration can be self-controlled. Unfortunately a recent attempt to replicate this finding failed (Brimer & Kamin, 1963).

Some of the work on pain perception in humans also supports this motivation hypothesis. Melzak (1961) mentions a study by Hill which demonstrated that *S*s perceive stimuli as less painful when they are told they can control the stimulus. And more recently, Lepanto, Monroney, and Zenhausern (1965) reported that pain thresholds are lower when *S*s cannot terminate the pain-producing stimulus.

Why escapable aversive events should provoke less "fear" (are less aversive) than inescapable ones has still not been satisfactorily determined. The cause may very well be just the self-control factor, or it may be the greater variability in intensity or duration of aversive stimulation that usually results from escape conditions. For example, in the present experiment, because of their own variable escape latencies the rats in the escape group had a more variable light duration than the rats in the "yoked" no-escape group whose punishment duration in a session was constant at the mean latency of the escape group.

REFERENCES

AZRIN, N. H., HAKE, D. F., HOLZ, W. C., & HUTCHINSON, R. R. Motivational aspects of escape from punishment. *J. exp. Anal. Behav.*, 1965, **8**, 31–44.

BRIMER, C. J., & KAMIN, L. J. Fear of the CS in avoidance training and fear from a sense of helplessness. *Canad. J. Psychol.*, 1963, **17**, 188–193.

DINSMOOR, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.*, 1954, **61**, 34–46.

DINSMOOR, J. A. Punishment: II. An interpretation of empirical findings. *Psychol. Rev.*, 1955, **62**, 96–104.

HEARST, E., & WHALEN, R. E. Facilitating effects of D-amphetamine on discriminated-avoidance performance. *J. comp. physiol. Psychol.*, 1963, **56**, 124–134.

HEFFERLINE, R. F. An experimental study of avoidance. *Genet. Psychol., Monogr.*, 1950, **42**, 231–334.

KELLER, F. S. Light aversion in the white rat. *Psychol. Rec.*, 1941, **4**, 235–250.

LEITENBERG, H. Response initiation and response termination: Analysis of effects of punishment and escape contingencies. *Psychol. Rep.*, 1965, **16**, 569–575.

LEPANTO, R., MONRONEY, W., & ZENHAUSERN, R. The contribution of anxiety to the laboratory investigation of pain. *Psychon. Sci.*, 1965, **3**, 475.

MALOTT, R. W., SIDLEY, N. A., & SCHOENFELD, W. N. Effects of separate and joint escape and avoidance contingencies. *Psychol. Rep.*, 1963, **13**, 367–371.

MARX, M. H., & HELLWIG, L. R. Acquisition and extinction of avoidance conditioning without escape responses. *J. comp. physiol. Psychol.*, 1964, **58**, 451–452.

MELZAK, R. The perception of pain. *Scient. American*, 1961, **204**, 41–49.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.

MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.

SOLOMON, R. L. Punishment. *Amer. Psychologist*, 1964, **19**, 239–253.

(Received July 16, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 400–405

FUNCTIONAL FIXEDNESS: STIMULUS EQUIVALENCE MEDIATED BY SEMANTIC-ACOUSTIC SIMILARITY[1]

SAM GLUCKSBERG AND JOSEPH H. DANKS

Princeton University

Ss were required to refer to a metal tool with E-assigned names while working on a problem in which that tool had to be used as a substitute for wire. Problem solving was facilitated when the assigned tool name rhymed with the word wires. When both the tool and wire were assigned nonsense-syllable names, no effect of rhyming was obtained, indicating that generalization along an acoustic semantic rather than a phonetic dimension mediated functional equivalence.

The functional fixedness paradigm, in which an object must be used in a novel way to solve a problem, offers an approach to the study of stimulus equivalence in the context of problem solving. When an object is used in a novel way, it may be considered equivalent, in certain respects, to the object for which it is substituted. For example, the solution to the circuit problem (Glucksberg, 1964) requires that *S* employ a metal tool as a substitute for wire in order to complete an electrical circuit. The metal tool may then be considered functionally equivalent to the wire it replaces.

If it is assumed that simple learning mechanisms are involved in complex problem-solving situations, then one should expect stimulus generalization to play a role in mediating this type of functional equivalence. In operational terms, problem-solving performance should be a function of the similarity between the two objects involved. The greater the similarity, the greater the likelihood that one object will be used as a substitute for the other. In a series of earlier studies (Danks, Glucksberg, & Weisberg, 1966), this was found not to be the case. Similarity of the color of a metal tool to the color of either wire insulation or of bare metal wire was unrelated to problem-solving performance.

One possible reason for this failure to obtain evidence for stimulus generalization in this context may be the similarity dimension employed. Thought has traditionally been viewed as a time-ordered or sequential process (Deese, 1966). If just this property of thought is considered, then audition would seem more closely analogous to thinking than is vision. Audition is time ordered, as is thought, while vision is essentially spatial and atemporal. This fairly superficial analogy gains credence with extension to language, which is also time ordered, particularly when we assume that acoustic variables are more intimately related to the character of language than are visual variables.

These considerations lead to the suggestion that similarity along an acoustic dimension may mediate functional equivalence where similarity along a visual dimension (i.e., hue) did not. Experiment I tests the hypothesis that availability of novel function is facili-

---

[1] This research was supported by Grants MH 10742-01 and MH 10742-02 from the National Institutes of Health, Public Health Service. This work made use of computer facilities supported in part by National Science Foundation Grant NSF-GP579. Theodore Feinstat's assistance in data collection is gratefully acknowledged.

tated by acoustic similarity between the name of the functionally fixed object and the name of the object for which it is substituted.

EXPERIMENT I

Method

Subjects.—One hundred and twenty male volunteers, all undergraduate students at Princeton University, served as paid *S*s.

Experimental task.—The *S* constructs two simple electrical circuits. The first serves as a training task and also provides a covariance control measure for individual differences in task-relevant skills. The second serves as a functional fixedness problem.

In the training task, *S* completes a series circuit between a 6-v. bulb and a battery, using metal binding posts to connect six lengths of wire. Enough wire is made available on request to complete the circuit, and a crescent wrench, pliers, or Channel-lock pliers is also available to loosen and tighten the binding-post nuts. These materials are kept out of sight behind a wood screen except when they are in use or have been placed into the circuit.

The functional fixedness problem requires *S* to complete a circuit consisting of two D cells, an SPST knife switch and a 6-v. bulb. The circuit components are mounted on an 8-in. × 12-in. rectangular Masonite pegboard. The D cells and bulb are prewired to four metal-binding posts (see Fig. 1). The *S* must complete the connections between the binding posts and nut-and-bolt terminals on the switch. Four lengths of wire are available, which, when combined, are 1½ in. too short to complete the circuit. A 6-in. wrench, pliers, or Channel-lock pliers (the same tool used in training), matched for gloss and hue, and a 12-in. coil of Plasticene modelling clay, ¼ in. in diameter, are also available. As in training, these materials are kept out of sight when not in use or in the circuit. The solution to the problem is to use the metal tool as a wire substitute.

Experimental design.—A 3 × 2 factorial design with three metal tools and two labels was employed, yielding six label-tool combinations. The 120 *S*s were randomly assigned to six experimental groups of 20 each. In three of these groups, *S*s referred to the metal tool as *pliers* and to the wire as *wires*. These three groups represented the high acoustic similarity condition. The other three groups referred to the tool as a *wrench* and to the wire as *wires*. These three groups represented the low acoustic similarity condition. All groups referred to the clay as *clay*.

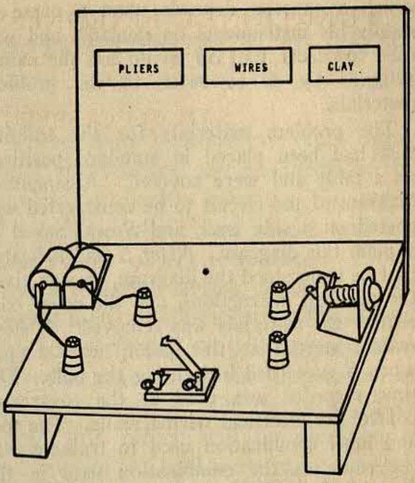


FIG. 1. Circuit problem employed. (High acoustic similarity condition is shown.)

Three different tools were used to control for possible effects attributable to the particular tool employed and for the possibility that assigning an incorrect name to a tool would facilitate problem solution by calling attention to the tool. Two groups, one in each similarity condition, were given a Channel-lock pliers, a tool that looks like a wrench. Either name, *wrench* or *pliers*, may be appropriate. Two groups had a wrench, referring to it either as a *wrench* or as *pliers*. The two other groups having a pair of pliers were similarly treated.

This design permitted isolation of effects attributable to labels, to the particular tool available, and to any interaction of tool and label.

Procedure.—In both the training and problem tasks, printed labels were mounted on the wooden screen affixed to the problem board (see Fig. 1). The *S*s were instructed to ask for the material they wished to use. They could keep any material they received so long as they were using it, and then had to return it to *E*, unless that material had been installed into the circuit. The *S*s were informed that they could ask for any item any number of times. If *S* asked for more wire after he had used up all the available wire, then he was told that no more wire was available. This procedure was adapted from the operating room procedures in which a surgeon depends upon a nurse to supply his instruments on demand, and was used to ensure that *S*s would use the names assigned by *E* to refer to the problem materials.

The problem materials for the training task had been placed in standard positions on a table and were covered. A simplified diagram of the circuit to be constructed was posted on a side wall, and *S* was asked to explain this diagram. After *S* had indicated that he understood the diagram, he was given the problem instructions, and the cover concealing the materials was removed. A stopwatch, started at this point, was stopped when *S* succeeded in lighting the bulb. The time recorded was used as the covariance control for electrical wiring skills. The tool and label combination used in training was the same as the combination used in the functional fixedness problem, which immediately followed the training task.

For the functional fixedness task *S* was again asked to explain a simplified diagram of the circuit to be completed, and was then given instructions. The measure of performance was time to solve, measured from the removal of the cover concealing the task materials until *S* had placed the metal tool into the circuit. Any *S* who failed to solve the problem within 20 min. was assigned a score of 20 min. and the experiment was terminated.

At the end of the experiment a printed questionnaire was given to all *S*s to determine (*a*) if *S*s could recall the materials that had been available in the verbal form designated by *E*, and (*b*) if *S*s would spontaneously verbalize any relation between the names of materials and the nature of the problem. The two questions were (*a*) What materials were available to you? and (*b*) What was the purpose of this experiment?

TABLE 1

Training Task Performance as a Function of Label-Tool Conditions

Available Tool	Label			
	Pliers		Wrench	
	M	*SD*	*M*	*SD*
Pliers	4.60	2.32	4.85	2.19
Channel-lock Pliers	4.80	2.20	4.78	1.87
Wrench	4.34	1.50	4.07	1.91

Note.—Measures in minutes.

TABLE 2

Problem-Solving Performance as a Function of Label-Tool Conditions

Available Tool	Label					
	Pliers			Wrench		
	M	*SD*	No. of Failures	*M*	*SD*	No. of Failures
Pliers	6.46	6.32	1	10.45	7.34	6
Channel-lock Pliers	9.09	6.90	2	14.25	6.99	9
Wrench	7.70	8.26	5	9.93	6.88	5

Note.—Measures in minutes.

Results and Discussion

Questionnaire.—All *S*s referred to the experimental materials in the verbal form designated by *E*. No *S* verbalized any relation between the sounds of the names used to refer to the tool and wires, nor did any *S* verbalize the purpose of the experiment.

Training.—All *S*s said that they understood the circuit diagram provided and each *S* could explain it satisfactorily. Every *S* succeeded in completing the training circuit. The means and standard deviations of the times to complete the training circuit are presented in Table 1. Analysis of variance applied to these data indicate no significant differences between groups as a function of label, tool, or interaction of label and tool, $F < 1.0$ for all contrasts. Analysis of variance of a reciprocal transform of the time scores yielded the same results.

Functional fixedness.—Means and standard deviations of solution times are presented in Table 2, together with the number of *S*s who failed to solve the problem within 20 min. Chi-square applied to the failure frequency data as a function of label (rhyming vs. nonrhyming) indicated that acoustic similarity did not reliably affect frequency of solution, χ^2 (1) = 1.72, $p > .05$, although the obtained difference is in the predicted direction.

A reciprocal transformation was applied to the solution time data, and analyses of variance were performed on both the transformed and original scores. Since the conclusions drawn from the results of these two analyses are identical, only the analysis of the transformed scores is reported. Problem-solving performance did not vary as a function of the particular tool available, F (2, 114) = 1.39, $p > .05$, nor was the interaction between label and tool a significant source of variance, F (2, 114) = .78, $p > .05$. These data indicate that the problem was not differentially difficult as a function of the tool supplied, and that referring to a tool with an incorrect name neither facilitated nor impaired problem-solving performance.

The predicted effect of acoustic similarity was obtained; problem-solving time was significantly shorter when the rhyming labels were used than when the nonrhyming labels were used, F (1, 114) = 11.11, $p < .01$. The labeling variable accounted for an estimated 14.5% of the total variance.

The same findings hold when training-task performance is taken into account. The Pearson product-moment correlation between reciprocal transforms of training time scores and problem-solution time scores was +.31 ($p < .001$). Analysis of covariance, using the reciprocal transform of training time scores to partial out task-relevant skills, was applied to the reciprocals of problem-solving time scores. Again, the effects of tools and of the interaction of tool and label were not significant. The main effect of labels was significant at the .01 level, F (1, 113) = 9.32. An analysis of covariance performed on the untransformed scores yielded the same conclusions.

These data demonstrate that functional equivalence can be mediated by acoustic similarity between the name of a functionally fixed object and the name of the object it replaces. However, a question arises concerning the specific locus of this effect. This effect may be attributed to primary stimulus generalization along the dimension of acoustic similarity. On the other hand, would this effect still be obtained if the name of the tool were to be acoustically similar to an arbitrary name assigned to the wire, a name which did not carry the semantic content of the word *wires?*

In other words, would acoustic similarity still facilitate functional equivalence when both the tool and the wire are given neutral names (such as nonsense syllables) which are acoustically similar but which do not denote conductive material? If primary stimulus generalization is indeed involved in mediating functional equivalence, then acoustic similarity between any arbitrarily designated names should facilitate problem solution. If, however, semantic content is a necessary condition for such facilitation, then the phenomenon observed in Exp. I is more akin to mediated or secondary generalization than it is to primary stimulus generalization. Experiment II tests the hypothesis that acoustic similarity of object names facilitates functional equivalence in the absence of semantic content.

EXPERIMENT II

Method

Subjects.—Forty male volunteers, all undergraduate students at Princeton University, served as paid *S*s.

Procedure.—The *S*s were randomly assigned to one of two groups of 20 each. One group used the labels PEEM and BEEM to refer to the metal tool and to the wires. Half of this group used PEEM to refer to the tool and BEEM to refer to the wires. The opposite condition held for the remaining *S*s in this group, which represented the high acoustic similarity condition. The low similarity group used the labels JOD and BEEM to refer to the tool and wires. Half the *S*s in this group referred to the tool as JOD, the other half as BEEM. All *S*s, irrespective of group assignment, referred to the clay as DAF. These nonsense labels were selected from those used by Dietze (1955).

The 6-in. crescent wrench employed in Exp. I served as the tool in both the training and problem tasks. The results of a separate study indicated that *S*s tend to use the metal tool as a conductor during pretraining when nonsense-syllable labels are employed (Glucksberg & Danks, in press). Therefore, the training task used in this experiment differed from that used in Exp. I in that the binding posts were 10 in. apart, precluding the use of the tool as a wire substitute. In all other respects, the procedures followed those of Exp. I.

Results

Questionnaire.—All *S*s referred to the experimental materials in the verbal form designated by *E*, and none verbalized the purpose of the experiment or commented on the similarity between tool and wire labels.

Training.—All *S*s said that they understood the circuit diagram and explained it satisfactorily. Every *S* succeeded in completing the training task. Mean times to complete the training task were 5.61 min. ($SD = 1.73$) and 5.33 min. ($SD = 2.15$) for the high and low similarity groups, respectively. These means are not significantly different.

Functional fixedness.—Three *S*s in the high, and four *S*s in the low similarity groups failed to solve the problem within 20 min. The mean solution times were 12.44 min. ($SD = 7.09$) and 12.48 min. ($SD = 7.40$) for the high and low similarity conditions, respectively. The median solution times for high and low similarity conditions were 13.18 and 13.35 min., respectively. Neither mean nor median solution times differed significantly.

Discussion

The data of Exp. II clearly indicate that mere acoustic similarity of names does not facilitate functional equivalence between two objects. The acoustic similarity of the nonsense names affected neither frequency of solution nor time to solve the problem. These data stand in contrast to those of Exp. I, where acoustic similarity of meaningful names did facilitate performance.

These findings are analogous to those obtained from studies of generalization of classically conditioned responses. Adult humans show greater generalization to stimulus items that are similar in meaning to the original training stimulus than to items that are similar in nonsemantic dimensions such as sound or visual appearance (Razran, 1961). Since semantic generalization is more effective than phonetic or phonetographic generalization in relatively simple conditioning situations, it should not be surprising that the same type of relationship holds in a more complex situation. If this analogy is more than superficial, then we should expect semantic similarity to facilitate functional equivalence in the absence of acoustic similarity. Unfortunately, data relevant to this point are lacking, but the experimental paradigm reported here would seem to offer rich opportunities for research directed at the specification of the component processes involved in complex problem solving.

The earlier failure to obtain evidence for stimulus-generalization effects along a visual dimension may, in the light of these considerations, be attributed to the apparent lack of any role played by primary stimulus generalization, irrespective of the sense modality or stimulus dimension involved. The appealing view of complex problem solving as a concatenation of simpler behavioral processes may still be useful, but it is clear that a simplistic application of S–R principles to such situations is inadequate. The question deserving further attention is not whether or not principles of conditioning and selective learning play a role in problem solving, but rather, which of these principles do in fact play a role, and under what sets of conditions. An adequate theoretical treatment of problem solving will have to account for both those cases where simple S–R principles do apply and those where they do not. In addition, it must specify those phenomena which require the postulation of principles not derivable from conditioning and selective learning. One step towards such a theory is the specification of the extent and limitations of the applicability of associationistic principles derived from simpler learning situations to problem solving.

REFERENCES

Danks, J. H., Glucksberg, S., & Weisberg, R. W. Functional fixedness: The role of stimulus generalization in problem solution. Paper read at Eastern Psychological Association meeting, New York, 1966.

Deese, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins Press, 1966.

Dietze, D. The facilitating effects of words on discrimination and generalization. *J. exp. Psychol.,* 1955, **50**, 255–260.

Glucksberg, S. Problem-solving: Response competition and the influence of drive. *Psychol. Rep.,* 1964, **15**, 939–942.

Glucksberg, S., & Danks, J. H. Effects of discriminative labels and of nonsense labels upon availability of novel function. *J. verbal Learn. verbal Behav.,* in press.

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.,* 1961, **68**, 81–147.

(Received July 18, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 406–413

EFFECT OF DIFFERENTIAL REINFORCEMENT UPON INFORMATION-GATHERING STRATEGIES IN DIAGNOSTIC PROBLEM SOLVING[1]

EDITH D. NEIMARK

Douglass College

6 experimental groups, differing with respect to the proportion (p) of training problems on which gambling would be rewarded (.375, .50, .625, .75, .875, 1.00) received 16 training problems. Data on number of moves to solution, strategy score, and proportion of 1st-move gambles were considered in relation to 2 theoretical models: a rational-decision model and a probability-matching model. The finding that both number of moves and strategy scores at the end of training are linear functions of p is compatible with predictions from both models. That proportion of 1st-move gambles is also a linear function of p is predicted only by the probability-matching model. However, finer examination of individual performance suggested that the data are better described by a rational-decision model and that Ss are more conservative than they theoretically should be.

Previous investigations of diagnostic problem solving (Neimark, 1961) have shown the development of an information-gathering strategy to be a function of the informational characteristics of a problem and of *S*'s prior history of reinforcement with respect to earlier problems. A number of response measures were shown to be simple linear functions of probability of reward for gambling (Neimark & Wagner, 1965). The present experiment attempts to clarify the nature of the effect of differential reinforcement by examining in greater detail its effect upon development of information-gathering strategies over an extended training series in relation to prediction from two theoretical models.

The structure of the diagnostic problem-solving task is illustrated in Fig. 1. At the outset *S* is given an answer sheet containing *n* patterns, or possible solutions, each composed of *m* binary elements. One of these patterns is concealed beneath a problem board with *m* movable shutters—one over each element of the pattern. The *S*'s task is to identify the concealed pattern by opening as few shutters as possible. Figure 1 shows four patterns, each composed of four black or white circles, and a board with one shutter open. The potential information available to *S* is completely determined by the composition of his answer sheet; his actual information-gathering behavior is characterized by a *strategy score* giving the mean expected informational outcome (in bits) of a series of shutter openings (moves). For example, if *S* began by opening the shutter at Positions A or D he would eliminate half of the patterns to be considered as possible solutions regardless of the color of the circle beneath the shutter. The expected informational outcome (reduction in uncertainty) of moves of this sort is 1 bit; such moves will be called *safe moves* and a series of safe moves will be called a *safe strategy.* If, on the other hand, *S* began by opening Shutter B or C, he would solve the

---

[1] This research was supported under Grant HD 1725-01 from the United States Public Health Service. The data were collected and analyzed by Sheila Feig. Carol Jorgensen also assisted in analysis of the data. The assistance of both is gratefully acknowledged.

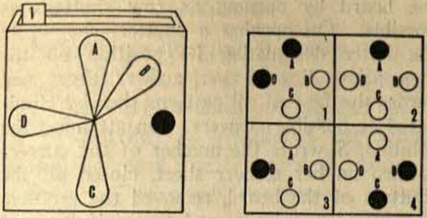


FIG. 1. Schematic of problem board with four shutters (Shutter B open) and an answer sheet with four patterns of four binary elements each. (Since the circle at B is black, only Pattern 4 can be the right answer. In this instance gambling is rewarded.)

problem in one move (get 2 bits of information) if Pattern 4 were concealed in the board, but would eliminate only one pattern (get .415 bits of information) if Pattern 1, 2, or 3 were concealed. If one assumes that patterns are selected randomly (i.e., the probability of concealment for each pattern is $1/n$),[2] then the expected informational outcome of opening Shutter B or C on the first move is .811 bits: $.811 = \frac{3}{4}(.415) + \frac{1}{4}(2.0)$. Moves of this sort will be called *gambles*. The strategy score is gotten by summing the expected informational outcome of each move over a series of moves and dividing by the number of moves. The maximum possible strategy score value of unity is obtainable only with a series of safe moves; it is lowered by gambling and the strategy score is reduced more by gambling on the first move than by gambles later in a series of moves.

In the experiment to be reported the answer sheet was so constructed that for each problem half the available first moves were gambles and the other half were safe moves. All gambles "paid off" on the same problem (as is the case in Fig. 1). The independent variable was the probability over a series of problems of reward for gambling, p, which was manipulated by varying the proportion of a set of problems on which gambling could yield a one-move solution. For example, if no patterns like 4 in Fig. 1 are ever concealed in the problem board, $p = 0$; if all the concealed patterns are like Pattern 4, then $p = 1.0$; if half the patterns concealed are like Pattern 4 and the other half are like Patterns 1, 2, or 3 then $p = .50$, etc. One group was continuously rewarded for gambling, one group was never rewarded for gambling, and the remaining groups were partially reinforced for gambling (on 3, 4, 5, 6, or 7 out of 8 problems) during their training series.

Results will be considered in relation to two a priori models for problem solving behavior: A *rational-decision model,* and a *probability-learning model* (a third possible a priori model, a *random-move model,* will not be considered since it does not describe the behavior of adults although it provides fairly accurate description of the performance of young children: Neimark & Lewis, 1967). The instruction to identify the concealed pattern by opening as *few* shutters as possible is uniquely satisfied by a simple rule: Always gamble if $p > .375$ and never gamble for $p < .375$.[3] Of course, S can intelligently apply the rule only if he knows the value of p, as he presumably does not at the outset of training. Thus, this rational-decision model would be expected to apply only for fully informed Ss, or at the end of a long training series. Another model, which allows for changes in performance over training series, is obtained by analogy with a probability-learning model for event prediction in the noncontingent case (Estes & Straughan, 1954). All available first moves may be classified on the basis of potential consequences into one of two mutually exclusive and exhaustive response classes: gambles and safe moves. If a gamble is successful then the E_1 operator is applied and, if it is not the E_2 operator applies. Reinforcement may be treated as noncontingent since at the end of each problem *S* must check his answer and see if it could have been obtained with fewer moves (such an assumption seems to be as tenable in this instance as in the case of a two-choice probability learning experiment). This model would lead to the prediction that at the end of a long training series probability of a first-move gamble should be a linear function of the probability of reward for gambling, p, whereas a rational decision model would predict that asymptotic probability of a first-move gamble should be a discontinuous function of p (going from zero for $p < .375$ to unity for $p > .375$).

---

[2] In practice, the probability of concealment of any given type of pattern can be any arbitrary value, p; and p is the independent variable of the present experiment. Strategy scores are, however, always computed on the assumption that $p = 1/n$.

[3] If S always begins by gambling then on a proportion, p, of the problems he will solve in one move whereas on the remaining problems, $(1 - p)$, four moves will be required. With a safe strategy all problems can be solved in exactly three moves. The expected number of moves to solution, $M = (p)(1) + (1 - p)(4) < 3$ for $p > .375$.

METHOD

Subjects.—Fifty-six girls from an introductory psychology course (who volunteered as part of a course requirement) were randomly assigned to seven groups of 8 *S*s each. Their modal age was 18.

Procedure.—The formal structure of the experimental task was very similar to the simplified version of Fig. 1. The *S* was given a 9-in. square wooden board with eight movable shutters equally spaced around a circle 8 in. in diameter, and a stack of eight dittoed answer sheets. Each answer sheet contained eight numbered patterns of black and white circles, one of which was presented in the problem board. The *S* was told to determine which of the eight patterns of the answer sheet was concealed in the board by opening as few shutters as possible. On opening a shutter, she wrote the letter designating its position on the appropriate line of her answer sheet and across the face of all patterns thereby eliminated as possible answers. On attaining the solution, *S* wrote the number of the correct pattern on her answer sheet, closed all the shutters of the board, removed the problem to verify her solution and to see if it could have been obtained in fewer moves. Four times during the course of the first eight problems *E* reminded *S* to see if she could have solved them in fewer moves. The *S* then went on to the next problem and its accompanying answer sheet. Eight problems, each in a tabbed pliofilm envelope, had been loaded into the board before the start of the experiment. On completion of the eighth problem, *S* was given a second set of eight problems. Nothing was said to indicate that the later set of problems differed in any way from the first. Each *S* worked at her own speed; time on each problem was obtained by means of a stopwatch.

Treatment of the seven experimental groups differed with respect to the proportion of the 16 problems which could be solved in one move: 0, .375, .50, .625, .75, .875, or 1.0. On completion of the last problem *S* was questioned as to her procedure for solving and what she thought was the purpose of the experiment.

Answer sheets were so designed that on each problem half of the available moves were safe moves and half were gambles (i.e., opening that shutter would yield 3 bits of information when gambling was rewarded and .2741 bits when it was not; its expected informational outcome was, therefore, .5436 bits). All four available gambling moves always "paid off" on the same pattern. Patterns were drawn in such a way that the one pattern on which gambling always "paid off" was not perceptually different from the rest. Within a set of eight problems each problem was done with a different answer sheet, the patterns of which were physically different in appearance from the patterns of the other answer sheets. Actually, all answer sheets were derived from the same prototype through permutation of (*a*) shutter positions within a pattern, and (*b*) arrangement of patterns on a sheet. Apparently no *S* realized this. The sequence of answer sheets was ordered by means of an 8×8 Latin square. For the second set of problems rows of the square were permutted. Thus, each *S* saw the same answer sheet twice in the course of 16 problems.

RESULTS

Before considering the adequacy of two models in describing the data, it is of interest to characterize the effect of proportion of problems on which gambling will be rewarded upon a number of response measures: moves, strategy scores, and solution time. With respect to the course of learning for each measure, a plot of group-mean learning curves over the 16 problems yields a family of curves which separate from the first problem on. Analysis of variance comparison of the seven experimental groups of the first eight problems shows that differences among groups are statistically significant at the .01 level or beyond; F (6, 49) = 33.380, 7.522, and 6.934 for moves, strategy scores, and time, respectively. For number of moves to solution, the learning curves are essentially flat and parallel as is indicated by a nonsignificant F ratio for both Trials and Trial × Group interaction. For strategy scores, on the other hand, there is a significant change in performance over trials and slope of group curves differs: for Trials F (7, 343) = 7.044, and F (42, 343) = 1.480 for Group × Trials interaction. The 0 group appears to be approaching a consistent safe strategy; the .75, .875, and 1.0 groups are approaching a gambling strategy; and the curves for the .375, .50, and .625 groups hover at an intermediate value (see Fig. 2). Mean time to solution is the only response measure which does not yield a consistent ordering among groups: The 1.0 group (which rapidly learns to solve in one move) solves in the shortest time whereas the .375 group seems to require the longest time throughout the series of 16 problems. The remaining groups vary considerably from one problem to the next but seem to be approaching roughly equivalent speeds on the final problems.

Each response measure seems to be

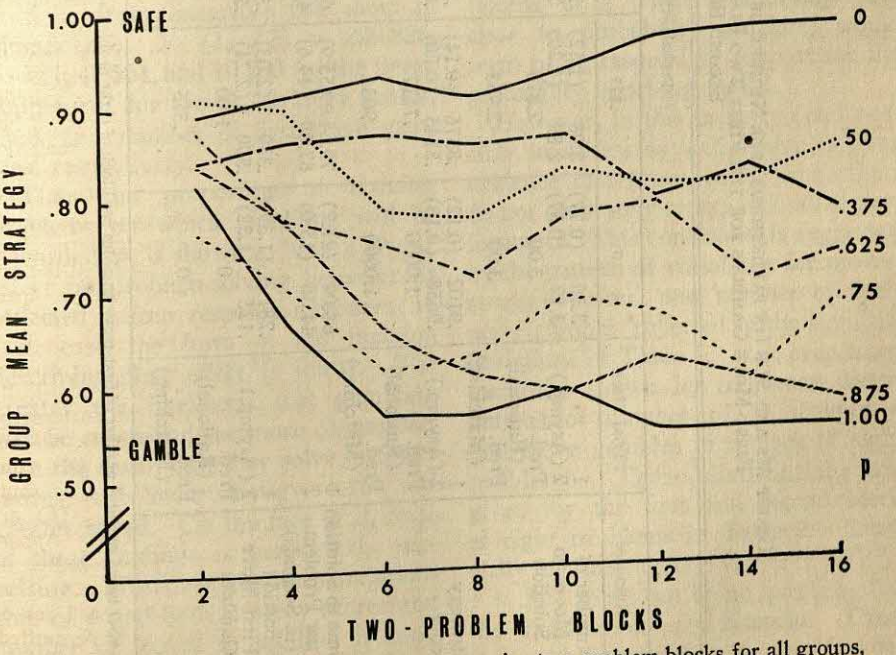


FIG. 2. Group mean strategy scores on successive two-problem blocks for all groups.

TABLE 1

Group *M*s and *SD*s (in parenthesis) for Number of Moves, Strategy Scores, Time to Solution, and Proportion of First-Move Gambles for the First Two Blocks of Eight Trials, along with Predicted Values for Second Block of Eight Trials

Response	Trial Block	Probability of Reward for Gambling: 0	.375	.50	.625	.75	.875	1.0
Number of Moves to Solution	1	3.50 (0.77)	2.98 (1.12)	2.70 (1.11)	2.44 (1.41)	2.11 (1.32)	1.78 (0.99)	1.39 (0.70)
	2 (Asymp.)	3.12 (0.57)	2.89 (1.20)	2.75 (1.22)	2.36 (1.35)	2.00 (1.52)	1.47 (1.07)	1.03 (0.18)
	Pred. (Rational)	3.00	2.875	2.50	2.125	1.750	1.375	1.00
	Pred. (Matching)	3.00	2.953	2.75	2.453	2.062	1.578	1.00
Strategy Score	1	.9102 (0.12)	.8576 (0.14)	.8413 (0.20)	.7772 (0.18)	.6998 (0.19)	.7216 (0.20)	.6532 (0.19)
	2 (Asymp.)	.9658 (0.13)	.8145 (0.20)	.8248 (0.21)	.7497 (0.20)	.6611 (0.17)	.5949 (0.16)	.5549 (0.10)
	Pred. (Rational)	1.0000	.7417	.7020	.6624	.6228	.5832	.5436
	Pred. (Matching)	1.0000	.9031	.8510	.7890	.7171	.6353	.5436
Time to Solution per Problem (in Seconds)	1	64.09 (21.2)	79.64 (55.9)	61.56 (27.8)	53.11 (33.1)	41.38 (30.8)	46.06 (29.2)	38.36 (28.1)
	2	44.36 (15.0)	67.39 (44.6)	48.59 (27.8)	42.88 (32.0)	32.50 (32.0)	36.95 (32.0)	20.42 (10.2)
Proportion First-Move Gambles	1	.281 (1.39)[a]	.359 (2.03)	.375 (2.78)	.547 (1.80)	.688 (3.84)	.656 (2.05)	.719 (.97)
	2 (Asymp.)	.125 (1.32)	.500 (2.24)	.500 (3.12)	.625 (2.24)	.781 (2.16)	.922 (1.65)	.969 (.044)
	Pred. (Rational)	0	1.00	1.00	1.00	1.00	1.00	1.00
	Pred. (Matching)	0	.375	.50	.625	.75	.875	1.00

[a] *SD* based on total number of first-move gambles per block of eight problems.

a simple linear function of *p*. To illustrate this, group means and standard deviations on successive eight-problem blocks are given for each measure in Table 1. In addition, predicted asymptotic values for number of moves and strategy scores have been given for each of the models under consideration. A fourth measure, proportion of first-move gambles is given in the bottom row of Table 1.

If one takes the mean of the last eight problems as an estimate of final level of performance (a procedure which seems justified in view of statistically nonsignificant *F* ratios for trials main effect) the between-groups main effect for the six groups given some reward for gambling may be subjected to orthogonal component analysis to determine the form of the function describing the effect of *p* upon each response measure. In each case the linear component is statistically significant at beyond the .001 level whereas the pooled remaining components (deviations from linearity) fall short of significance: $F\ (1,\ 42) = 299.826$, 53.721, 41.561, and 19.200 for the linear component for moves, strategy scores, time, and number of first-move gambles, respectively.

Thus the proportion of training problems on which gambling will be rewarded, *p*, is shown to have a strong effect on problem-solving behavior as reflected in four response measures; in each case the form of the function describing that effect is linear. The greater the likelihood that gambling will be successful the more *S*s gamble, with the result that they solve in fewer moves, with lower strategy scores, and greater speed. On the face of it, none of these findings is particularly surprising. Moreover, the linear relationship of *p* and both strategy scores and number of moves to solution is compatible with predictions from both of the models to be considered (in actuality, the predicted relation for the probability-matching model has a slight curvature, but probably not enough to be detected by orthogonal component comparisons). Interestingly enough, the obtained values—in the case of both number of moves and strategy scores—fall between the theoretically predicted values (where slightly higher values are predicted from the probability-matching model than from the rational-decision model). Clearly, neither measure provides a decisive test between the two models.

With respect to proportion of first-move gambles, on the other hand, although both models yield the same prediction for the 1.0 and 0 groups, they differ markedly in predictions for the remaining groups: The probability-matching model predicts a linear relation of proportion of first-move gambles and *p*, whereas the rational decision model predicts a discontinuous function which breaks at $p > .375$. Our finding of a linear relationship very close to probability matching would seem to be resounding support for the probability-matching model.

However, in this instance (and possibly in others as well) there is clear evidence that group mean proportions do not accurately reflect individual performance. This conclusion is suggested by the pattern of variability for moves, strategy scores, and number of first-move gambles reflected in the standard deviations of Table 1. It is even more forcefully shown by frequency distributions of number of *S*s making *n* first-move gambles in a block of eight problems. These distributions are given for the first and second block of eight problems in Table 2. Especially for intermediate values of *p* (e.g., $p = .50$) there can be no question that the distributions are bimodal. Thus, it would appear that *S*s tend to fix

TABLE 2

Number of *S*s in each Group Making *n* First-Move Gambles on Eight Successive Problems

Group p	Problems 1–8									Problems 9–16								
	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
0	1	2	1	2	2					4	2	1		1				
.375	1	1	1	2	2			1		1	1		1	1	2	1		1
.500	3		2	1				2		3						3	2	
.625		1	1		1	2	3					1	2	1	1		1	2
.750		1		1			3	2	1			1		1	1		1	4
.875		1		1			5		1				1					7
1.000					1	2	3	2									2	6

Note.—Frequencies are given for Problems 1–8 and 9–16.

upon one of two consistent rules:[4] Always gamble on the first move, or always follow a safe strategy; and that it is the proportion of *S*s in a group who adopt a gambling strategy, rather than the proportion of first-move gambles for each *S*, which reflects probability matching. Why this should be the case is not at all clear from the present data; that it is, in fact, true is further indicated by the result of post-experimental queries summarized in Table 3.

Only two *S*s reported anything even vaguely resembling a probability-matching rule: One *S* in the 0 group reported gambling occasionally "when she got bored," and one *S* in the .375 group said occasionally she "took a chance" when she "felt lucky." The vast majority of *S*s reported gambling (and the number of these increases consistently as probability of reward for gambling increases) or a safe strategy (and their number decreases with p). In addition, one *S* reported a position rule (always start with A), which is fairly common among children (Neimark & Lewis, 1967), and even among adults at the outset of a problem series. Two *S*s, both in the .75 group, reported a strategy of gambling on the *second* move—a strategy which, in fact, they consistently employed. A second-move gamble strategy, while irrational, is commonly found in the data of the present study and of earlier ones (e.g., Neimark & Wagner, 1965) and may be regarded as evidence of conservatism.

Discussion

We have shown that the effect of the proportion, p, of problems on which one-move solutions are obtainable has a powerful effect upon information-gathering behavior in diagnostic problem solving. For four response measures (moves, strategy scores, solution time, and proportion of first-move gambles) the nature of this effect is accurately described by a

TABLE 3

Number of *S*s Reporting Each of a Variety of Strategies on the 16 Training Problems

Strategy	Group						
	0	.375	.50	.625	.75	.875	1.0
Safe	6	3	3	1			
Gamble	1	4	5	6	6	8	8
Vary	1	1					
Other				1[a]	2[b]		

[a] Fixed position.
[b] Second move gamble.

---

[4] The measurement of consistency, and of increasing stereotypy, is tricky business. Uncertainty analyses (Newman, 1951) have been done for the data of the present experiment and will be sent to anyone who volunteers to explain what they mean.

linear function. The impressive orderliness of this effect, however, does little to clarify the mechanisms which produce it. To this end, two theoretical models were considered; one an analogue of probability matching in a simple two-choice task, the second a rational decision model. Although the proportion of first-move gambles as a function of *p* was in accord with predictions from probability-matching and incompatible with predictions from a rational-decision model, examination of the frequency distributions of first-move gambles led to rejection of the probability-matching model. Postexperimental reports, and comparison of those reports with obtained strategy scores, on the other hand, clearly showed that (*a*) *S*s formulate a rule and (*b*) their strategy is accurately described by the rule. The results of not only the present experiment but many others (e.g., Neimark & Lewis, 1967) have convinced the author that in this situation, with increasing mental age, it is the case that *S*s do increasingly formulate and follow a general rule. Once behavior is brought under control of a rule, trial to trial variations in the structure of the problem or of reinforcement contingencies have relatively little effect. The more consistent the experimental conditions of reinforcement, the more quickly *S* can formulate a rule; and the more dramatic the change in experimental conditions, the more readily *S* comes to recognize that a change in rule is in order. Changes in rule, when they occur, may very well take place in an all-or-none fashion. If so, then one might expect that some version of an all-or-none pattern model (Atkinson & Estes, 1965) might well provide accurate description of information gathering behavior—certainly a simple probability-learning analogy does not.

Although *S*s do seem to be developing rational strategies, details of their performance are not accurately described by the rational-decision model which was considered. Specifically, although *S*s do tend to adopt either a consistent safe strategy or a consistent gambling strategy, they do not seem to shift to the former from the latter at $p > .375$ but at $p > .50$. This suggests that our data are best described by a rational-decision model with a correction for conservatism. There is independent evidence for the existence of conservatism among college students in decision-making tasks not only in the present study but also in e.g., Phillips, Hays, and Edwards (1966). That college students tend to be more conservative than other *S* groups is suggested by an unpublished comparison of the performance of New York University students and Air Force recruits run with eight problems and varying *p* values in which AF *S*s showed much more gambling than college students. One speculative interpretation of this difference among individuals might be that formulating and following a general principle is more reinforcing in and of itself—especially for intellectually inclined *S*s—than obtaining limited heuristic strategies. Just what variables operate to produce conservatism is a question beyond the scope of the present experiment.

REFERENCES

Atkinson, R. C., & Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology,* New York: Wiley, 1963. Pp. 121–268.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.,* 1954, **47,** 225–234.

Neimark, E. D. Information-gathering in diagnostic problem-solving: A preliminary report. *Psychol. Rec.,* 1961, **11,** 243–248.

Neimark, E. D., & Lewis, N. The development of logical problem-solving strategies. *Child Develpm.,* 1967, **38,** in press.

Neimark, E. D., & Wagner, H. Effect of reward conditions upon information gathering in diagnostic problem solving. *Mimeo Tech. Rep.,* 1965.

Newman, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.,* 1951, **64,** 252–262.

Phillips, L. D., Hays, W. L., & Edwards, W. Conservatism in complex probabilistic inference. *IEEE Trans. Hum. Factors Electron.,* 1966, **HFE-7,** 7–18.

(Received July 18, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 414–419

AN ANALYSIS OF CUES IN SERIAL LEARNING[1]

WILLIAM H. SAUFLEY, Jr.

Emory University

2 methods of serial presentation were used to investigate the serial-position hypothesis and the specificity hypothesis. In 1 method, the 1st item is variable from trial to trial and in the other, constant. Ss learned an 11-item list to 1 perfect recitation under 1 of the 2 methods and then relearned the same list either by the same method or the other method. 2 lists were used, one having strong item-to-item associations, the other weak item-to-item associations. The variable 1st-item method and the weak-association list were significantly harder to learn. Transfer was higher with the strong-association list and no change in presentation method and least with the weak-association list and change in method from constant to variable 1st-item method. The results were interpreted as supporting the serial-position hypothesis while showing the utility of sequential or item-to-item associations in learning and transfer.

Several recent papers have described the status of research on the functional stimulus in serial learning (Battig, Brown, & Schild, 1964; Jensen & Rohwer, 1965; Underwood, 1963). The two most straightforward hypotheses about serial learning identify the functional stimulus as either (*a*) the immediately preceding item, called the chaining hypothesis, or (*b*) the position an item occupies in the list, called the serial-position hypothesis (Young, 1961). In investigating the role of these two possible sources of cues in learning a serial list, it is not the intent of this paper to insist that there is *one* stimulus in serial learning; rather, it is to discover what cues are predominant as various independent variables are manipulated.

The second of the two hypotheses above, the serial-position hypothesis, derives from characteristics of presentation of a serial list. Namely, each trial is begun with the same first item, and ends with the same last item, and all other items hold a constant position as required by the sequential order of the list. A presentation technique has been used which does not employ constant serial or *E*-given position (Bowman & Thurlow, 1963; Ebenholtz, 1963; Keppel, 1964; McManis, 1965; Winnick & Dornbush, 1963). The procedure consists of presenting a different item from the list as the first item on each trial, then to proceed through the entire list with each of the items in sequence. Since this latter procedure is analogous to spinning a memory drum in between trials and randomly picking a first item, for convenience it will be referred to as the spin technique. Accordingly, the customary serial-presentation method will be referred to as the set technique.

If *S*s tend to use the constant serial position of an item for a cue, as suggested by the serial-position hypothesis, then the spin condition would, and does, according to the studies just cited, increase markedly the difficulty in learning. Following the same line of reasoning, if a list is learned with position constant, the set condition, and then the task is changed to the spin condition, a decrement in performance should result due to the removal of position cues. The degree of sequential association that takes place should determine the amount of transfer. On the other hand, if learning is first by the spin procedure and the switch is to the set condition, the decrement should be negligible since the list would have been learned by sequential associations and the addition of position cues should not interfere with giving the correct responses. Accordingly, the present design uses both spin to set and set to spin transfer conditions.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements of the PhD degree at Northwestern University. The author gratefully acknowledges the aid and encouragement of Benton J. Underwood.

It is assumed that learning under the spin condition cannot take place by learning to *E*-given position cues but rather is dependent upon the formation of sequential associations. A procedure, then, which would increase the ease of formation of sequential associations would reduce the number of trials required to master the list. In order to provide such a test, a list was used in which the associations between successive words in the list were high. Another list was also used in which the associations between successive words were low. Learning the list with strong sequential associations was expected to be more rapid, particularly under the spin procedure, where such associations would be required for learning the list.

METHOD

Design and Ss.—Two levels of association between successive words in the list, two methods of presentation during the first stage, and two methods of list presentation during the second stage were combined to yield a 2 × 2 × 2 factorial design and eight independent groups.

A total of 176 *S*s were drawn from the introductory psychology students at Northwestern University with the only restriction being that they not have served in other serial-learning experiments. The *S*s were assigned randomly upon appearance at the laboratory to one of the eight conditions to make up groups of 22 *S*s each.

Materials.—Two lists were constructed using common English words equated for frequency of occurrence. One list, the high association list, was constructed so as to make sequential or chaining associations likely. The word sequence was CLUB, HIT, SOCK, FOOT, SHOE, SHINE, BRIGHT, LIGHT, DARK, FOREST, and WOOD. Two words in the list, CLUB and SOCK, served a double meaning in order to make sequential associations continuous. The low-association list sequence was BRIGHT, WOOD, SKILL, CARROT, HIT, DRINK, GRADES, HOME, FOREST, SPORT, FOOD. Not all of the words used in these particular lists appear in association norms so the lists were presented to 39 *S*s who were asked to rate the lists on the degree to which each word led to the one which followed. Each list was rated as a whole on a 9-point scale. Mean ratings on the high- and low-association lists were 7.2 and 1.6, respectively, the higher number indicating greater association.

Procedure.—Each *S* was assigned to either the high- or the low-association list and remained with that list throughout the two stages of the experiment. Stage 1 and Stage 2 learning were both carried to a criterion of one perfect recitation. After the completion of Stage 1, transfer was made to the appropriate Stage 2 list-presentation condition, half of the *S*s continuing with the same method and the other half with the changed method.

In the Spin condition items are presented in the same sequence but serial positions vary from trial to trial. Since each word could appear as the initial item, a sequence of 11 "lists" was constructed in which each word appeared as the start word in a randomly determined manner. The same sequence of "lists" was used for all *S*s within a given level of association, and each list appeared equally often on the first trial. Again, 11 lists were constructed for the set condition so that each of the 11 words occurred as the initial item in the list. Since there were 22 *S*s in each group, a particular word appeared as the initial item in the list for 2 *S*s within a given condition, the same holding true for the spin method. Under the set condition, of course, the initial word remained the same on each trial whereas under the spin condition it varied from trial to trial.

In transfer, on Stage 2, the start word in transfer for the groups which changed presentation method was determined by the start word on the criterion trial. For example, when transfer was from spin to set, the start word on the set condition was the same start word that would have been given on the criterion-plus-one trial.

It should be noted that learning an identical list under the spin and set conditions requires a different number of sequential associations to be formed. Specifically, since the start word is not consistent in the spin condition, an association must be formed between each successive word, including those occupying the first and last position or any given trial. For an 11-item list as used here, 11 adjacent associations could be formed. The arrangement under the set procedure would provide for only 10 such associations. Thus, learning by the spin procedure required one more sequential association than set. In order to compensate for this inequity, a set of three asterisks was used as the cue item in set. In transfer, this meant that an *S* going from spin to set would have to learn to give the first word in the list in response to the asterisks whereas in going from set to spin, he would have to learn to associate two words not previously appearing sequentially, the first and the last.

Complete information was given in each stage of the experiment in an effort to minimize the variability which could be produced by *S*s "catching on" to the procedure at different rates. The lists were presented at a 2-sec. rate with a 2-sec. intertrial interval. The time between completion of Stage 1 and the start of Stage 2 was approximately 80 sec.

Results and Discussion

Stage 1.—List H, the high-association list, was learned more rapidly under set than spin, 4.98 and 8.95 trials, respectively. List L, the low-association list, was more difficult to master than List H, but again fewer trials were taken under set than spin, 12.84 and 32.64, respectively. Heterogeneity of variance was found, $F_{\max}$ (43) = 20.88, $k = 4$, $p < .01$, but the data were not transformed. Both list construction and method of presentation produced significant effects. List L was significantly more difficult than List H ($df = 1{,}172$ for all *F*s in Stage 1) $F = 133.60$, $p < .01$, and the spin condition more difficult than set, $F = 76.13$, $p < .01$.

Furthermore, it was intended that the arrangement of words in List H would facilitate the formation of sequential associations particularly with the spin procedure. The interaction of method and list was significant, $F = 33.54$, $p < .01$. The larger difference between the two lists under spin points to the increased utility of sequential associations when constant *E*-given position is not available as a cue.

Backward errors, or repetition of items that have appeared earlier in the trial, are usually small in number when set procedure is used. An evaluation was made to determine if the removal of the position cue, as in the spin procedure, affected backward errors. Expressed in terms of the percentage of total errors, backward errors under set in Stage 1 were about 4.5% while under spin they were about 16.5%. Many of these errors occurred in the position of the first item anticipated on a given trial for *S*s learning under spin. That is, if Word n appeared as the first word, the response would be Word $n - 1$ instead of $n + 1$. The percentage of times that these backward errors occurred in the first position ranged from 20% with List L to 76% for List H. Thus the high proportion of backward errors obtained with the Spin procedure were due in great part to errors made on the very first item anticipated on a given trial and this relationship particularly held for List H.

A serial-position analysis of the spin method was made in which the most frequently correct response for each *S* was placed in the first serial position. The remainder of the words were then placed in successive positions. Curves were found which resembled those obtained by Lippman and Denny (1964) and Breckenridge, Hakes, and Young

(1965). These curves have been interpreted as being due to idiosyncratic primacy cues or anchor points established by *S*s in the absence of *E*-given serial position.

Stage 2.—The four basic groups from Stage 1 were each split into two groups in Stage 2. One-half continued under the same method of presentation while the other changed to the alternate method. For convenience the groups will be referred to by the procedures used in the two stages of the experiment. A designation of set-spin would then refer to a group which learned a list following the set procedure in Stage 1 and then relearned the same list following the spin procedure in Stage 2. Statistical tests were made to compare each of the four pairs of groups to determine whether the groups differed in Stage 1 learning. All of the *F*s were less than one and it was concluded that the two subgroups from each of the four major groups were equivalent prior to Stage 2 treatment.

Figure 1A shows the mean trials required to reach a criterion of one perfect recitation the second time by the groups learning List H. In Fig. 1B the same data are presented for List L. The groups continuing under spin or switching to spin took more trials than the comparable groups continuing under set or switching to set. The spin procedure in Stage 2 required more trials to master than the set procedure ($df = 1, 168$ for all *F*s in Stage 2), $F = 21.51$, $p < .01$, and the groups learning List L took longer than the groups learning List H, $F = 19.21$, $p < .01$. The training method (Stage 1) also produced a significant effect in Stage 2, $F = 7.56$, $p < .01$, but its effect becomes more apparent when considering the interactions.

It was expected that the switch in Stage 2 would produce different effects

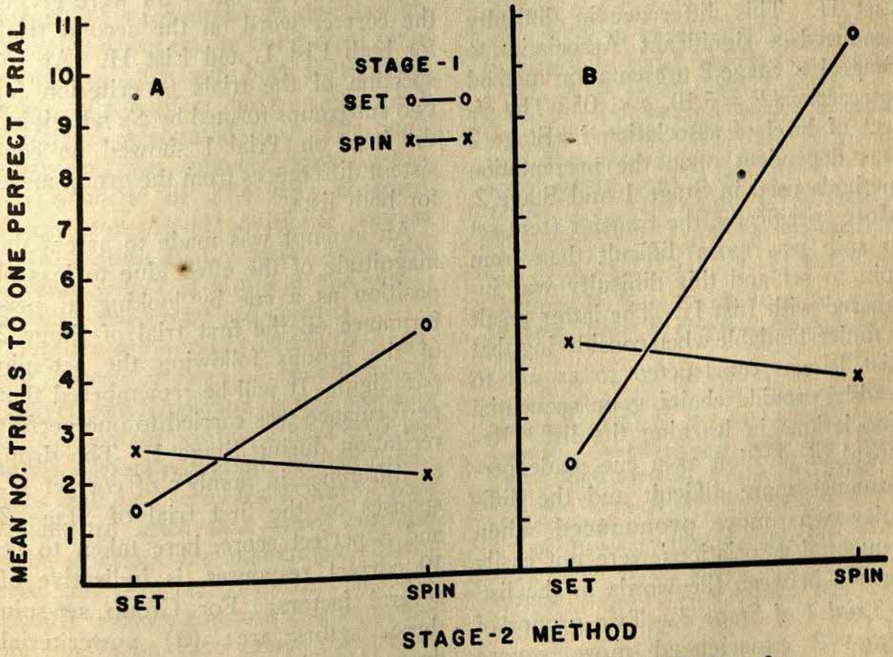


FIG. 1. Mean number of trials to one perfect recitation in Stage 2 for List H (2a) and List L (2b).

depending upon whether the switch was from spin to set or set to spin, the former requiring fewer trials to mastery than the latter. The Stage 1 × Stage 2 interaction was significant, $F = 37.40$, $p < .01$. The groups learning under the spin-set paradigm took less than one trial more to learn the list again under set than the groups continuing with the spin procedure. This was true even though the words in the list now maintained the same serial position from trial to trial. The List × Stage 2 interaction was also significant, $F = 4.73$, $p < .05$. This interaction is indicated in Fig. 1 by the greater number of trials required in Stage 2 under the spin method, particularly with List L.

Lists also produced differences in the number of trials taken to master the list when transfer was from set to spin. The number of trials required to learn under the set-spin condition was markedly greater with List L than with List H. This difference in difficulty produced a significant Association × Stage 1 × Stage 2 presentation method interaction, $F = 5.10$, $p < .05$. The effect of level of association in Stage 2 was dependent upon the presentation methods used in Stage 1 and Stage 2. More specifically, the transfer from set to spin was more difficult than from spin to set and this difficulty was increased with List L. The latter result is understandable when considering that List L was constructed so as not to readily provide chaining or sequential associations in learning the list. Removal of position as a cue made performance more difficult and the difficulty was more pronounced when sequential associations were not easily formed between the words in the list.

Trial 1 of Stage 2.—Two groups of Stage 2 experienced more severe changes than the other groups, namely those learning under the set-spin paradigm. Specifically, the first word changed on each trial in Stage 2 and the asterisks were removed; the cue for the previously first word was now the previously last word. It was possible that some of the difficulty exhibited by these two groups could have been due to failure to respond when the first word had to be anticipated. A tally was made of the number of *S*s responding correctly and the number who did not make a response, that is, the number of *S*s who anticipated the first word from Stage 1 correctly and those who did not. (The tally indicated that no incorrect responses were given.) For the groups learning List H, 15 *S*s had the correct response and the remaining seven did not anticipate. Of the 22 *S*s learning List L, 12 anticipated correctly while 10 failed to respond. Responding or failure to respond was not dependent upon the list used, $\chi^2 = .86$. Additionally, all but 4 *S*s were giving the correct word on the second trial, for both List L and List H. An inspection of the trials to criterion for two subgroups formed by *S*s not giving the item on Trial 1 showed no consistent differences from the group mean for both lists.

An attempt was made to assess the magnitude of the effect due to loss of position as a cue by looking at performance on the first trial of Stage 2 of the groups following the set-spin paradigm. It will be remembered that performance was carried to one perfect recitation during Stage 1. The drop, or difference in number of correct responses on the first trial of Stage 2 and a perfect score, here taken to be 10 correct responses, is indicative of several factors. For Group set-spin these factors are: (*a*) postcriterial drop; (*b*) change of method; (*c*) removal of position cues. To illustrate the procedure followed to eliminate Factors *a* and *b*, data from List H will be used. The postcriterial drop observed following training on set was assessed by the mean drop in set-set, here .50 items. To obtain an estimate of the second effect, a change in conditions from Stage 1 to Stage 2, the drop score from Group spin-set, 1.27 items, was obtained. However, in order to account for the greater postcriterial drop in this measure resulting from slower learning under spin, the drop score from spin-spin, .86 words, was subtracted from 1.27 to leave a net of .41 words representing the effect due to a change in conditions, over and above the postcriterial drop. When the two scores representing postcriterial drop and the effect due to change in conditions, .50 and .41 words, respectively, are summed, a score of .91 items results. When this in turn is subtracted from the drop score in set-spin, 1.73 a difference of .82 items remains. This remainder represents Factor *c*, the effect due to removal of serial position as a cue with both postcriterial drop and the effect due to a change in conditions accounted for. The same procedure was followed with List L and a value of 2.54 was obtained. These scores lend additional support to the notion that a lesser drop was suffered with List H due to the availability of sequential associations.

Apparently serial position is a potent source of cues. The absence of *E*-given serial position during learning and removal of position after a criterion of one perfect recitation, yielded conditions making a serial list of words more difficult to learn than when constant serial position was maintained. However, the very fact that the lists were learned without constant *E*-given position indicates that other cues were available, particularly where List H was learned more rapidly than List L. Some evidence for anchor points was found in the analysis of the spin method in Stage 1, and this evidence is consistent with other research (Breckenridge et al., 1965; Lippman & Denny, 1964). Although the curve obtained here resembled the customary bowed serial-position curve, it is not sufficient for inferring that *S* constructs serial-position cues throughout the list.

REFERENCES

BATTIG, W. F., BROWN, S. C., & SCHILD. Serial position and sequential associations in serial learning. *J. exp. Psychol.*, 1964, **67**, 449–457.

BOWMAN, R. E., & THURLOW, W. R. Determinants of the effect of position in serial learning. *Amer. J. Psychol.*, 1963, **76**, 436–445.

BRECKENRIDGE, K., HAKES, D. T., & YOUNG, R. K. Serial learning in a continuous serial list. *Psychon. Sci.*, 1965, **3**, 139–140.

EBENHOLTZ, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *J. exp. Psychol.*, 1963, **65**, 603–608.

JENSEN, A. R., & ROHWER, W. D., JR. What is learned in serial learning? *J. verbal Learn. verbal Behav.*, 1965, **4**, 62–72.

KEPPEL, G. Retroactive inhibition of serial lists as a function of the presence or absence of positional cues. *J. verbal Learn. verbal Behav.*, 1964, **3**, 511–517.

LIPPMAN, L. J., & DENNY, M. R. Serial position effect as a function of intertrial interval. *J. verbal Learn. verbal Behav.*, 1964, **3**, 496–501.

MCMANIS, D. L. Position-cues in serial learning. *Amer. J. Psychol.*, 1965, **78**, 668–671.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 33–48.

WINNICK, W. A., & DORNBUSH, R. L. Role of positional cues in serial rote learning. *J. exp. Psychol.*, 1963, **66**, 419–421.

YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–518.

(Received July 20, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 420–424

RESPONSE STRENGTH AS A FUNCTION OF PRE- AND POST-REWARD DELAY AND PHYSICAL CONFINEMENT

ROBERT L. WILLIAMS

West Georgia College

The experiment was conducted to appraise the utility of reinforcement and contiguity explications of nonchaining delay of reward learning. Of particular relevance was the role of incompatible behavior occurring during the delay intervals. The relationship of 3 delay periods (prereward, no delay, and postreward), 2 goal-box forms (restrictive and unrestrictive), and trial blocks to initial response speed and running speed was assessed. A measurement was also made of magnitude of activity during pre- and postreward delays in the large goal section. 60 male hooded rats were used as Ss in the 6 experimental groups. The results showed that prereward delay is decremental to performance ($p < .001$) but postreward delay is not ($p > .05$). There was no significant difference in the activity levels of pre- and postreward delay Ss.

The present study attempted to empirically compare reinforcement and contiguity interpretations of nonchaining delay of reward learning by focusing on the role of incompatible behavior which occurs during the delay interval. Roberts (1930) noted that introducing a delay between the instrumental response and reward resulted in a diverting of the animals' responses to features of the situation not immediately relevant and consequently led to a prolonging of the learning period. Contiguity theorists such as Spence (1956), Guthrie (1935), and Estes (1950) would propose that if these nongoal-oriented responses occur to essentially the same stimulus components as does the instrumental response, they likewise become conditioned to the stimulus complex and impede performance. However, since these competing responses are followed by reward, reinforcement formulations (Hull, 1951; Miller, 1951) would likewise imply that they are learned.

Perhaps a more decisive way to differentiate the two views is to focus on the effect of a delay imposed after reward. If one could demonstrate that the stimulus complex during postreward delay is comparable to that associated with the instrumental response and that competing responses actually do occur in postreward-delay intervals, contiguity theory would apparently predict less proficient performance under postreward delay than with no delay. In contrast, since these incompatible responses are not followed by reward, reinforcement explications would not predict a performance decrement with postreward delay.

Previous studies have demonstrated that postreward delay is not comparable to prereward delay. In fact, Davis (1954) and Fehrer (1956) have shown that postreward delay *S*s perform at essentially the same level as no delay animals. However, there is some indication in these studies that incompatible behavior did not occur to a marked degree in the postdetention period. The apparatus dimensions listed by *E*s indicate that *S*s were confined in rather restrictive goal sections. Carlton (1954) found that a prereward delay of 10 sec. has a more detrimental effect on performance in a large goal box than in a confining one, presumably because of the greater opportunity to engage in nongoal-oriented response in the large box. In addition, Fehrer noted a tendency for postreward delay animals to be less active during delay than prereward. However, she reported no objective measurement of incompatible behavior.

To supplement previous postdetention findings, *E* studied the relationship of prereward delay, no delay, and postreward delay to performance and the effect on behavior of imposing such delays in restrictive and unrestrictive goal sections.

METHOD

Subjects.—The *S*s were 60 male hooded rats, approximately 100–110 days old at the beginning of the study. They were assigned without bias to six equal groups.

Apparatus.—The apparatus consisted of three compartments (starting box, runway, and goal box) separated by two guillotine-type doors. The starting box was 6 in. in length, the runway 24, and both were 3 in. wide and 3 in. high. A restrictive and unrestrictive goal box were used, the restrictive being 6 in. × 6 in. × 6 in. and the unrestrictive 6 in. × 18 in. × 12 in. (see Fig. 1). The apparatus was painted aluminum and covered with wire mesh. The restrictive goal section was made of clear Plexiglas and could be inserted into the large goal box. The reward mechanism consisted of a drinking cup presented 1 in. above the floor of the goal box through a ½-in. hole in a metal indentation in the rear wall of the goal section. The circular indentation was 2¼ in. in diameter and 1 in. in depth. Three-tenths milliliter of 20% sucrose solution was used as reward for all groups.

Immediately after *S* was placed in the starting box and achieved orientation toward the runway, Guillotine Doors 1 and 2 (Fig. 1) were raised. Both doors were held open by 6-v. ac electromagnets. To assess latency in leaving the starting box, a photoelectric cell was installed 6 in. from Door 1. When the animal intercepted the light beam, the photoelectric cell was energized and Door 1 released, thus rendering the return path to starting box inaccessible. After the rat had entered the goal section, Door 2 was electronically closed by *E*, thus prohibiting *S* from returning to the runway.

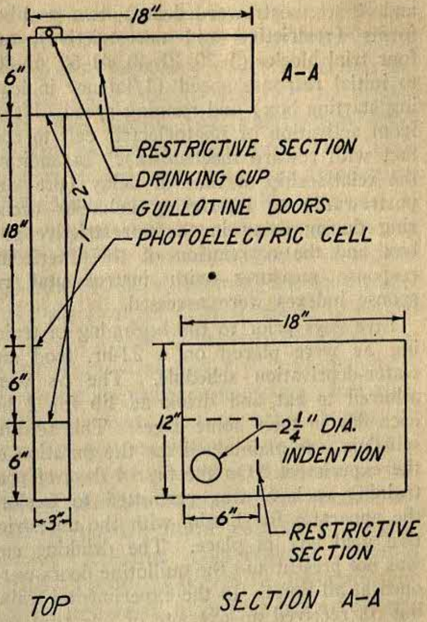


FIG. 1. Plan of the apparatus.

Two Standard clock timers were used to assess latency and running time. Both timers were activated by the raising of Guillotine Door 1. Timer 1 was stopped when *S* energized the photoelectric cell. Stopping of Timer 2 occurred when the high input impedance, gas tube switch was short circuited, which in turn was caused by the animal's simultaneous contact with any part of the metal indentation and the floor of the goal box (Dawson, Lindner, & Perry, 1963). The short circuiting invariably occurred when *S*s attempted to approach the drinking cup.

To assess extent of movement during the delay interval, a phonograph crystal was attached to the underside of an aluminum floor mounted on springs. As the rat moved in the goal section, the floor would vibrate slightly. These vibrations were converted into electrical signals by the phonograph crystal and transmitted to a high voltage gain amplifier. The electrical signals were then recorded by an Esterline Angus strip-chart recorder.

Procedure.—An orthogonal design was used to assess the relationship of three delay periods (30 sec. prereward delay, 0 delay, and 30 sec. postreward delay), two goal-box forms (restrictive and unrestrictive), and four trial blocks (1–20, 21–40, 41–60, 61–80) to initial response speed (1/latency in leaving starting box) and running speed (1/time from activation of photoelectric cell to contact with reward mechanism). In addition, the relationship of time of delay (pre- and postreward) to interfering behavior occurring during delay in the unrestrictive goal box and the correlation of the interfering response measures with instrumental response indexes were assessed.

Five days prior to the beginning of training *S*s were placed on a 22-hr. food and water-deprivation schedule. The *S*s were allowed to eat and drink ad lib for 2 hr. each day in their home cages. This feeding schedule was maintained for the duration of the experiment. On the first 4 days of pretraining each *S* was permitted to explore the apparatus for 1 min. with the unrestrictive goal box in place. The drinking cup was not present and the quillotine doors were open at all times. In the experimental training *S*s received an average of one trial per day the first 10 days, two trials a day for the next 10 days, three trials per day for the third 10-day interval, and four trials a day for the last 5 days of the experiment. The daily trials were preceded by at least 12 hr. deprivation and spaced by 1½ hr. One hour following the completion of the daily trials the animals were fed and watered in the home cages.

The animals of Group 30-R received a 30 sec. prereward delay in the restrictive goal box and *S*s of Group 30-U a 30 sec. prereward delay in the unrestrictive. Group R-0 was run to the confining goal section with no delay and U-0 to the large goal box with no delay. Group R-30 received a 30 sec. postreward delay in the small goal box and U-30 an equivalent delay in the large goal section.

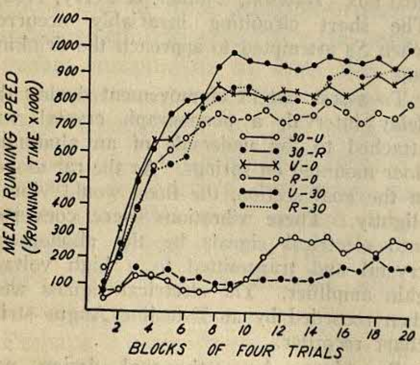


FIG. 2. Mean running speed for each group computed by blocks of four trials.

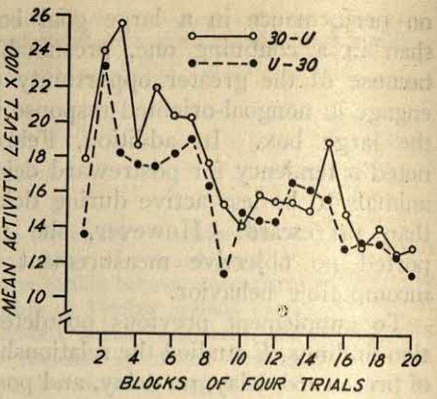


FIG. 3. Mean activity level for Groups 30-U and U-30 computed by blocks of four trials.

RESULTS

The analysis of initial-speed and running-speed scores indicated that prereward delay is detrimental to instrumental behavior, but postreward delay is not (see Fig. 2). There was no significant difference between the no-delay and postreward delay groups ($p > .05$), but both had significantly faster starting speeds and running speeds than the prereward delay *S*s, $p < .001$. Goal-box form was not significantly related to either of the instrumental response indexes.

The *E* was also able to objectively assess the number and extent of the animals' movements during the delay intervals. Figure 3 demonstrates that both groups were considerably more active in the early trials than later in training. Overall, there was not a significant difference in the activity levels for Groups 30-U and U-30. The relationship between incompatible behavior and performance level was further studied by correlating mean activity levels of the individual animals over the 80 trials with the mean initial response speeds and mean running speeds over the 80 trials. No significant correlation was obtained.

DISCUSSION

The results of the study appear to provide considerable support for a reinforcement interpretation of nonchaining delay of reward conditioning. Incompatible behavior occurring after reward did not affect performance decrementally. It is more difficult to evaluate the implications of the findings with respect to contiguity formulations. To provide a fair test of the contiguity paradigm, it must be demonstrated that the postdetention stimulus complex is comparable to that associated with the instrumental response and that nongoal-oriented behavior occurs in the postreward delay interval. The latter requirement was confirmed by the comparison of the activity levels of Groups 30-U and U-30. However, there would likely be considerable diversity of opinion concerning the extent to which the present study satisfied the former requirement.

Contiguity theorists emphasize the change in the stimulus complex as a result of reward. According to Guthrie (1935) reward adds nothing to the associative strength of the S-R bond, but rather protects it from disintegration by removing the animal from the stimuli associated with the instrumental response. In the context of Guthrie's problem box, the escape physically removed the organism from the prereward stimuli. But in the present experiment the postreward *S*s remained in the same physical situation as that accompanying the instrumental response. Therefore, whatever change that occurred in the stimulus complex had to be within the organism.

Both Guthrie and Spence have emphasized the importance of internal cues such as maintaining stimuli and drive stimuli (S_D) in comprising part of the stimulus complex. To what extent were internal cues altered by the conditions of the present study? Three-tenths milliliter of sucrose solution could hardly be construed as leading to complete cessation of all drive stimuli for *S*s having experienced at least 12 hr. food and water deprivation. One prime indicator of drive-stimuli intensity is the activity level of the organism. The finding that postreward delay *S*s were as active as the prereward lends support to the contention that drive stimuli had not been appreciably altered. It must be noted that reinforcement theory (Hull, 1950) does not dictate complete reduction of drive stimuli, but only minimal reduction for reinforcement to occur.

Since it would be extremely difficult to demonstrate that any two stimulus situations are entirely synonymous, most conditioning must involve some degree of stimulus generalization. According to a contiguity interpretation the greater the reward, the less should be the generalization of postdetention responses to prereward stimulus components. Slough (1960) found, however, a tendency for postreward delay to have a more deleterious effect with large reward than with small. The present *E* contends that, while the stimulus complex of postreward delay is not completely equivalent to that of prereward, by virtue of stimulus generalization contiguity theory should predict less proficient performance with postreward delay than with no delay. The finding that postreward and no delay *S*s performed at comparable levels plus the finding that response strength was not related to the size of the goal section raises crucial questions concerning contiguity explanations of nonchaining delay of reward learning.

REFERENCES

CARLTON, P. L. Response strength as a function of delay of reward and physical confinement. Unpublished master's dissertation, State University of Iowa, 1954.

DAVIS, A. D. A test of one aspect of contiguity theory. *J. exp. Psychol.*, 1954, **48**, 275–277.

DAWSON, W. W., LINDNER, R. S., & PERRY, N. The electrical quantification of the consumption of liquids by rats. *Amer. J. Psychol.*, 1963, **76**, 687–689.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

FEHRER, E. Effects of amount of reinforcement and pre- and postreinforcement delays on learning and extinction. *J. exp. Psychol.,* 1956, **52**, 167–176.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.,* 1950, **57**, 173–180.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.

ROBERTS, W. H. The effect of delayed feeding on white rats in a problem cage. *Pedagogical Seminary,* 1930, **37**, 35–38.

SLOUGH, D. A. Duration of reinforcement and delay after reinforcement in the runway. *Dissertation Abstr.,* 1960, **20**, 3382–3385.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received July 21, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 425–430

INTERLIST CHANGES IN SUBJECTIVE ORGANIZATION DURING FREE-RECALL LEARNING[1]

ANTHONY J. MAYHEW

University of Vermont

Several measures were used to analyze practice effects during free-recall learning. 24 Ss learned 6, 30-word lists of multisyllabic nouns. Each S experienced individually 10 consecutive learning and test trials on each of 2 lists per experimental session for 3 consecutive days. The Ss were equally divided into 2 groups on the basis of differential instructions. Conclusions are: (a) interlist increases in subjective organization and list learning occur within and between experimental sessions and vary as a function of differential instructions; (b) intertrial retention increases concomitantly with list learning, while intratrial retention remains invariant over lists and between conditions; (c) a tendency to repeat stimulus-list words more than once in recall decreases as a function of lists practiced.

Typical of several verbal learning tasks is the finding that when *S*s learn a series of unrelated lists in the laboratory, there is a noted improvement in list learning with practice (Postman & Schwartz, 1964). Using the method of free-recall learning, for instance, an interlist improvement in learning has been observed by both Dallett (1963) and Tulving, McNulty, and Ozier (1965). In addition, Tulving et al. (1965) found that there was an improvement in the subjective organization (SO) of the recalled words by *S*s over lists learned. The present experiment was designed to look at the relationship between interlist changes in subjective organization and learning during free-recall learning. The measure, subjective organization, developed by Tulving (1962b), is defined in terms of *S*s' tendency to recall items in the same order on different test trials in the absence of any experimentally manipulated sequential organization among items in the stimulus list. List learning (C), the measure of changes in learning over lists, is defined as the number of correctly recalled words per trial, without regard for order of recall.

In the attempt to assess the relationship between interlist changes in subjective organization and learning, different instructions were given to two groups of *S*s, independent of changes in learning materials or conditions of practice (cf. Tulving, 1962a). If, under these conditions, differences in subjective organization and learning are observed between the groups, it is possible that a dependent relationship may exist between the two variables. There were two other procedural details, the reasons for which should be noted. First, the learning situation was extended over several days so as to measure changes in the two variables, not only within but also *between* experimental sessions (cf. Thune, 1951). Second, a multitrial procedure for each list learned was used to maximize interlist practice effects (Dallett, 1963; Tulving et al., 1965).

---

[1] A portion of this paper was reported at the Eastern Psychological Association at New York City, April 1966. The work was carried out during the author's tenure as a postdoctoral fellow at the Institute of Human Learning, University of California, Berkeley. Appreciation is expressed to Endel Tulving for helpful suggestions in the design and data analysis of the experiment.

In addition, refinements in the data analysis were undertaken using two measures, intratrial retention (NC) and intertrial retention (CC), developed and described by Tulving (1964), for trial-by-trial analysis of the protocols of *S*s learning to free recall a list of unrelated words. Intratrial retention is defined as the number of correctly recalled words on Trial *n* which were *not* correctly recalled on Trial $n-1$. Intertrial retention is defined as the number of correctly recalled words on Trial *n* which *were* correctly recalled on Trial $n-1$. Intratrial retention and intertrial retention are, therefore, two logical and additive components of the number of correct responses per trial.

METHOD

Experimental design.—The *S*s learned to free recall six different lists of unrelated words. Each *S* experienced two lists during each daily experimental session for 3 consecutive days. During the learning of each list, *S* experienced 10 learning trials alternating with 10 test trials. There were two experimental conditions. Two groups of *S*s learned the same materials under identical conditions of presentation and recall, but with different sets of instructions. One group was given standard free-recall instructions, Cond. S, and the other was told, in addition, to organize their recalled words during test trials, Cond. O.

The experimental design is a three-factor experiment, 2 × (2 × 3), with repeated measures on the last two factors. The first factor is the independent main effect of instructions on Cond. S and O, and the last two factors are the main effects of lists learned, within and between experimental sessions, respectively.

Materials.—The learning materials were nouns selected from the Thorndike-Lorge (1944) word book, two or three syllables in length, and occurring at least once per million words in frequency count. These words were used to construct six, 30-word lists. In constructing these lists the words were selected so that each list contained words (*a*) approximately distributed among all frequency-of-occurrence categories, (*b*) with rough alphabetical distribution of the first letter of words, and (*c*) with no two words beginning with the same two-letter sequence of the alphabet.

From each of the six, 30-word, stimulus lists, 10 orders were generated by (*a*) initially randomizing the arrangement of the words in each list 10 times, (*b*) systematically reducing the frequency of repeated word pairs and triplets among the orders, and (*c*) separating any highly associated pairs of words in each order. Each set of 10 orders was presented as the 10 learning trials of an experimental list. All 60 lists, 10 for each of the six experimental lists, were recorded verbally on magnetic tape, using a Wollensak tape recorder. The words on each list were read into the tape at the rate of approximately one word per second. Longer intervals of time on the tape were allowed between lists.

Subjects.—The *S*s were 24 undergraduates of both male and female sex from introductory classes at the University of California, Berkeley. Twelve *S*s were assigned to each condition in the following manner. The first six consecutive *S*s were assigned to Cond. S, the next six consecutive *S*s were assigned to Cond. O, and the last 12 *S*s were evenly distributed between both conditions, but the schedule of assignment of *S*s to conditions was alternating rather than consecutive. Apparently, no *S* bias was introduced by this method of assignment since the differences of the means on all dependent variables between the first and last *S*s in each condition were negligible. The six experimental lists were counterbalanced in order of presentation to form six orders. This gave two counterbalanced list orders for each group of 12 *S*s, the same list order being assigned to two *S*s in each condition.

Procedure.—The *S*s participated individually in the experiment. At the beginning of the first experimental session, each *S* was instructed by *E* in the nature of the task. The *S*s in Cond. S were told that they would hear a series of word lists presented via loudspeaker and, that following each list, they were to say aloud as many words as possible, regardless of order. In addition to the standard free-recall instructions of Cond. S, *S*s in Cond. O were told to try to organize their recalled words in some manner on test trials, and this request was repeated at the beginning of each experimental session. Methods for organizing recall, suggested in the instructions, were (*a*) the grouping of words according to similarities, images, or associations, (*b*) relating words sequentially

TABLE 1

Mean Scores on all Measures and Correlations between Subjective Organization Scores and List-Learning Scores as a Function of Experimental Lists and Conditions

Days	Lists	Measures													
		Cond. S							Cond. 0						
		SO	C	SO:C	CC	NC^1	NC^2	RLW	SO	C	SO:C	CC	NC^1	NC^2	RLW
1	1	.210	16.73	+.85***	11.05	6.18	5.18	1.43	.235	18.73	+.61*	13.42	6.50	4.12	1.30
	2	.242	17.58	+.64*	11.96	6.44	4.81	1.32	.338	20.60	+.44	15.72	6.47	3.29	0.83
2	3	.221	16.98	+.74**	11.55	6.10	4.77	1.05	.310	18.81	+.56*	13.83	6.27	3.70	0.93
	4	.269	17.73	+.84***	12.33	6.38	4.42	0.86	.373	20.62	+.70**	15.83	6.37	3.21	0.72
3	5	.220	17.55	+.70**	12.11	6.29	4.58	0.86	.366	20.42	+.75**	15.43	6.57	3.41	0.58
	6	.285	17.91	+.79**	12.24	6.67	4.67	0.58	.420	20.97	+.58*	16.17	6.44	3.16	0.46

Note.—Means based on $N = 12$ per cell, NC = C − CC, NC^1 = Trials 1–5, NC^2 = Trials 6–10.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

by means of alphabetization, or making up stories among items, and (*c*) any other organizational methods which might be obvious to *S*. After the instructions each *S* learned two experimental lists. During learning trials, *S* listened to the auditory presentation of the words from a tape recorder. On test trials, *S* said aloud as many words as he was able to recall in one minute, while *E* recorded them by writing down the first two letters of each word. A second tape recorder, which also recorded the recalled words of *S*, provided a later check if *E* was unsure of some aspect of his record. Ten seconds elapsed between the end of a test trial and the beginning of a learning trial. One minute elapsed between the last test trial of one experimental list and the first learning trial of the next experimental list. Each *S* learned two lists per day and returned 24 and 48 hr. later for two more experimental sessions. Each experimental session lasted approximately 1 hr.

Results and Discussion

For each recall test of each *S*, the number of correctly recalled words was determined without respect to order in recall. Scores for each list on each of the measures, subjective organization (SO, Lag O), list learning (C), intratrial retention (NC), and intertrial retention (CC), were calculated on a 7096 IBM computer, using programs developed by Tulving. Each score was based on the total block of test trials per list for each *S*. Each *S*, therefore, has six scores on each measure, one for each of the six experimental lists learned. For each experimental list, mean scores on each of the measures are displayed in Table 1. The means are based on the 12 *S*s within a condition. Analyses of variance were performed on the scores of each of the measures.

Subjective organization and performance.—The interlist increases in SO scores both within and between experimental sessions were very significant, $F(1, 22) = 40.79$, $p < .01$, and $F(2, 44) = 9.53$, $p < .01$, respectively. The interlist increase in SO scores within experimental sessions replicates the finding by Tulving et al. (1965). The very significant interlist increase in C scores within experimental sessions, $F(1, 22) = 9.90$, $p < .01$, replicates the observations by Dallett (1963) and Tulving et al. (1965). However, the interlist increase in C scores between experimental sessions fell short of significance, $F(2, 44) = 2.70$, $p > .05$, and this may be attributable to the small number of *S*s in the conditions.

Within each condition, a product-moment correlation was computed between SO and C scores for each experimental list. The correlations are displayed in Table 1. The correlations are all positive and with the exception of one case, significant. No orderly relations appear in the table. Similar high correlations have been reported elsewhere (Tulving, 1962b, 1964).

Instructions.—The high positive correlations between SO and C scores and the concomitant changes between the two variables as a function of lists learned, suggest that the two may be related. The hypothesis is supported by the differential effect of instructions on the variables. First, SO scores were significantly greater in Cond. O than Cond. S, $F\ (1, 22) = 6.40$, $p < .05$, and there was a significant interaction between instructions and lists practiced, $F\ (2, 44) = 3.68$, $p < .05$, i.e., SO scores in Cond. O increased at a greater rate as a function of lists learned than did SO scores in Cond. S. Second, while the difference fell short of significance, C scores were consistently greater in Cond. O than in Cond. S, $F\ (1, 22) = 2.97$, $p > .10$. The results suggest that the instructions to organize increased both SO scores and C scores. A similar relationship has been reported by Tulving (1962a), in the case of *S*s learning to free recall a single list of unrelated words, namely that intralist increases in learning and subjective organization varied concomitantly as a function of differential instruction. The results of this experiment, then, reinforce the argument by Tulving (1962a) and Tulving et al. (1965) that list learning is dependent on the growth of subjective organization.

Intratrial and intertrial retention.—The interlist increase in CC scores within experimental sessions was very significant, $F\ (1, 22) = 9.35$, $p < .01$, while the interlist increase between experimental sessions was not significant, $F\ (2, 44) = 2.30$, $p > .10$. Neither of the interlist increases in NC scores, within or between experimental sessions, was significant, $F\ (1, 22) = 2.04$, $p > .10$ and $F\ (2, 44) = 1.46$, $p > .10$, respectively. The significant interlist increase in CC scores within, and nonsignificant interlist increase between, experimental sessions parallels the interlist changes in C scores. On the other hand, NC scores, both within and between experimental sessions, remain relatively constant as a function of trials, although a small downward trend may be detected.

Instructions.—The CC scores were greater in Cond. O than Cond. S, but the difference fell short of significance, $F\ (1, 22) = 4.02$, $p > .05$. Thus, the effect of instructions on CC scores was similar to the effect on C scores. The NC scores in Cond. O were significantly *less* than the scores in Cond. S, $F\ (1, 22) = 10.66$, $p < .01$, but this can be shown to be a procedural artifact involving a ceiling effect in list learning. Tulving (1964) has shown that NC scores rapidly decrease as *S* reaches learning asymptote, due to the fact that the number of new words remaining to be learned becomes very restricted. In the present experiment, the greater the C scores during list learning, the greater the possibility that the learning asymptote was reached. Hence, it follows that the finding of lower NC scores in Cond. O than Cond. S is expected, since the C scores are greater in Cond. O. Moreover, the differences in NC scores between the conditions should be negligible on earlier test trials, since the possibility of a ceiling in learning being reached is not likely. This prediction was tested by separating the NC scores for each experimental list into two blocks of trials, earlier learning (Trials 1–5), and later learning (Trials 6–10). Summarized in Table 1 are the mean NC scores for Trials 1–5 and Trials 6–10, as a function of lists learned and conditions. The prediction is supported by the findings that in earlier learning, Trials 1–5, there is no difference in NC scores between Cond. S and O, $F\ (1,\ 22) = 0.054$, $p > .25$, whereas in later learning, Trials 6–10, the NC scores in Cond. O are significantly less than the NC scores in Cond. S, $F\ (1,\ 22) = 7.40$, $p < .05$. It is concluded that list-by-list increases in C scores, as a function of list learned or instructions, were caused by increases in CC scores, while NC scores remained invariant (cf. Tulving, 1964).

Repetitions of list words.—Extralist intrusions were isolated, found to be negligible, and were not analyzed. However, a small number of "repetitions of list words" (RLW), occurred, i.e., stimulus-list words correctly recalled more than once within a given test trial, and since there appeared to be an orderly decrease in them as a function of lists learned, they were analyzed. Summarized in Table 1 are the mean RLW scores as a function of lists learned and conditions. The interlist increases in RLW scores, both within and between experimental sessions, were very significant, $F\ (1,\ 22) = 11.39$, $p < .01$ and $F\ (2,\ 44) = 10.93$, $p < .01$, respectively. The RLW scores were greater in Cond. S than Cond. O, but the difference was not significant, $F\ (1,\ 22) = 0.628$, $p > .25$. Thus, in free-recall learning, *S*s are able both to increase the number of correctly recalled words on a given test trial as a function of lists learned, and at the same time, to decrease a tendency to repeat stimulus-list words in recall.

It may be concluded that when *S*s learn to free recall different lists of words each day for several consecutive days, there is an improvement in list learning both within each daily experimental session and over days. This learning improvement was indicated by an increase in the average number of correctly recalled words per list, and a decrement in the tendency for *S*s to repeat stimulus-list words more than once in recall. This increase in the number of correctly recalled words per list is caused by an increase in the ability of *S*s on any given trial to remember words recalled correctly on the previous trial, i.e., intertrial retention. On the other hand, the number of new words which can be learned on any given trial remains invariant, i.e., intratrial retention. Application of the measure, subjective organization, revealed that *S*s increase their sequential order of recalled words both within and between experimental sessions. Finally, the fact that subjective organization scores and the number of correctly recalled words per list were highly correlated and, moreover, differential instructions to two groups of *S*s produced concomitant changes in both subjective organization and list learning, suggests the possibility that these two variables are dependently related.

REFERENCES

Dallett, K. M. Practice effects in free and ordered recall. *J. exp. Psychol.*, 1963, **66**, 65–71.

Postman, L., & Schwartz, M. Studies of learning to learn: I. Transfer as a function of method of practice and class of verbal materials. *J. verbal Learn. verbal Behav.*, 1964, **3**, 37–49.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Thune, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, **42**, 250–256.

Tulving, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canad. J. Psychol.*, 1962, **16**, 185–191. (a)

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–354. (b)

Tulving, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychol. Rev.*, 1964, **71**, 219-237.

Tulving, E., McNulty, J. A., & Ozier, M. Vividness of words and learning to learn in free recall learning. *Canad. J. Psychol.*, 1965, **19**, 242-252.

(Received July 22, 1966)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

The titles and authors of accepted papers are listed here. In 1965 this program was being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology and the outcome of this trial is presently being evaluated.

This listing represents the entire backlog of manuscripts accepted by this journal. They appear here 2 or 3 months after acceptance due to the printing lag. Such listing is intended to allow readers to become aware of research many months in advance of journal publication.

Effect of Initial Nonrewarded Trials: Factors Responsible for Increased Resistance to Extinction: Norman E. Spear* and Joseph H. Spitzner: Department of Psychology, Rutgers University, New Brunswick, New Jersey 08903.

Cue Utilization in Serial Learning: Sam C. Brown* and Eugene D. Rubin: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Strength of Alternative Responses and Subsequent Choices: William B. Pavlik* and Robert Lehr: Department of Psychology, Rutgers University, New Brunswick, New Jersey 08903.

An Investigation of Conditions Determining Contrast Effects in Differential Reward Conditioning: H. Wayne Ludvigson*: Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.

Operant Conditioning of Spontaneous GSRs: Negative Results: Robert M. Stern*: Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

Some Experiments on Simple and Choice Reaction Time: Joan Gay Snodgrass, R. Duncan Luce, and Eugene Galanter*: Department of Social Psychology, Columbia University, New York, New York 10027.

Proactive Inhibition and Competition at Recall: John P. Houston*: Department of Psychology, University of California, Los Angeles, California 90024.

Variables Affecting the Performance of Preschool Children in Intradimensional, Reversal, and Extradimensional Shifts: Corinne C. Mumbauer* and Richard D. Odom: Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

Function of Test Trials in Paired-Associate Learning: Chizuko Izawa*: Department of Psychology, San Diego State College, San Diego, California 92115.

Repeated Shifts in Reward Magnitude: Evidence in Favor of an Associational and Absolute (Noncontextual) Interpretation: E. J. Capaldi* and David Lynch: Department of Psychology, Mezes Hall 211, University of Texas, Austin, Texas 78712.

Transfer of Differential Eyelid Conditioning Through Successive Discriminations: David A. Grant,* C. Michael Levy, Johanna Thompson, Craig W. Hickok, and Dennis C. Bunde: Department of Psychology, Charter at Johnson, University of Wisconsin, Madison, Wisconsin 53706.

Anagram Solving as a Function of Bigram Rank and Word Frequency: Roger L. Dominowski*: Department of Psychology, University of Illinois at Chicago Circle, Box 4348, Chicago, Illinois 60680.

Extinguishing Avoidance Responses as a Function of Delayed Warning Signal Termination: Richard Katzev*: Department of Psychology, University of California, Santa Barbara, California 93106.

(Continued on page 442)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 431–442

HEART RATE DURING CONDITIONING IN HUMANS: EFFECT OF VARYING THE INTERSTIMULUS (CS-UCS) INTERVAL[1]

SHIRLEY E. HASTINGS AND PAUL A. OBRIST

School of Medicine, University of North Carolina

The interstimulus interval (ISI) was varied, using .8, 7, and 13 sec., in a test of whether the monophasic deceleration found during heart-rate conditioning of humans is a CR developed through CS-UCS pairing, an orienting response, or a previously conditioned response mediated by anticipation. The deceleration was small and early with an .8-sec. ISI and progressively larger and later with longer ones, favoring the last hypothesis. Analysis indicates that respiration did not cause the deceleration. A conditioned GSR was larger with the .8-sec. interval than with either of the others, suggesting that the sympathetic acceleration found previously under atropine may also be augmented by the short interval.

The present study investigates a unique effect of heart-rate conditioning in humans: the CR is often reported to be opposite in direction to the UCR. Although the UCR is usually acceleration, recent investigations by Obrist and others indicate that the CR is a monophasic deceleration when rate and depth of respiration are controlled throughout the experiment (Wood & Obrist, 1964) and that this is caused by activation of the vagus but is not merely a homeostatic response to a rise in blood pressure (Obrist, Wood, & Perez-Reyes, 1965). These findings present a major problem of interpretation since they do not fit the predictions of either Pavlov's stimulus-substitution theory or Cannon's theory of sympathetic arousal.

Reports of heart-rate deceleration contradicting one or both theories have actually appeared for many years. Sherrington (1900) noted that a dog decelerated from 150 to 54 beats/min on hearing a noise previously associated with shock. Although this deceleration may have been unusually large because of an earlier high spinal transection, it is clear evidence of vagal activation in an aversive situation and seems to fit neither theory. Others (Bechterev, 1928; Kling, 1933) have maintained, contrary to Cannon, that the vagus normally dominates during fear. Kling suggested that Cannon's theory holds only for fear involving fight or flight. He argued that cats are angered rather than frightened by a barking dog (Cannon's fear-producing stimulus) and stated that they showed deceleration when placed in a small enclosure or strapped on their backs, presumably more frightening situations. Several *E*s have reported evidence, often contrary to Cannon, of a deceleratory CR in both humans and other animals not only when the UCS was deceleratory (Kosupkin & Olmstead, 1943; Kotharevsky, 1936; Ssorokhtin, Turgel, & Minut-Ssorokhtin,

---

[1] This investigation was supported by Research Grant MH 07995 from the National Institute of Health, United States Public Health Service, to the second author. A portion of this paper was read at the Society for Psychophysiological Research Meeting, Houston, 1965, and a portion was submitted by the senior author to the University of North Carolina as a thesis in partial fulfillment of the requirements of the Master's degree, 1966. Appreciation is due Norma Thomas and Diane Little for their technical assistance.

1934), but also, contrary to Pavlov, in some *S*s and situations when the UCS was acceleratory (Beier, 1940; Black, 1965; Black, Carlson, & Solomon, 1962; Gantt, 1944; Peters & Gantt, 1951; Reese & Dykman, 1960; Robinson & Gantt, 1947).

Interest in the problem was slight, however, until Notterman, Schoenfeld, and Bersh (1952) demonstrated an apparent deceleratory CR clearly and consistently in humans and suggested that it was due to operationally defined "anxiety" (Schoenfeld, 1950). Many investigations have appeared since then of which two are particularly relevant to the present study. Smith and Zeaman, 1964, and Zeaman and Smith, 1965, have tried to determine whether the biphasic response they found with beat-by-beat analysis is explainable by stimulus substitution, drive reduction, or mediation theory, or is due to respiratory influence on heart rate. Their findings suggest both mediation by attention and respiratory influence, but their control of respiration, which consisted of sustained inspiration or expiration after CS onset only, was both aphysiological and diverting. Moreover, Obrist and his associates (Obrist et al., 1965; Wood & Obrist, 1964) have demonstrated that respiratory changes do not cause the deceleration since *S*s trained to maintain normal rate and depth of breathing throughout conditioning show a monophasic deceleration.

The present experiment tests several suggested hypotheses by varying the interstimulus interval (ISI), using .8[2] and 13 sec. as well as the 7 sec. previously used. First, it has been suggested that the deceleration is a CR developed because of CS-UCS pairing but inexplicably opposite in direction to the UCR. If this is true, it should be larger with a more optimal .8-sec. ISI and smaller with a 13 sec. one, as other CRs have typically been (Kimble, 1961). Also, if it is a CR, this experiment tests whether it is a CS- or UCS-bound CR. With a 7-sec. ISI, it appears about 5 sec. post-CS onset (2 sec. pre-UCS onset) and peaks on test trials about 9 sec. post-CS onset (2 sec. post-UCS onset). It may therefore be either a CS-bound CR with a long latency, appearing near UCS onset only because of the 7-sec. ISI, or a UCS-bound CR that comes to anticipate the UCS slightly. If it is CS-bound, its latency should be unaffected by changes in the ISI, but if it is UCS-bound, it should always appear near UCS onset.

If the results indicate that the deceleration is not a CR developed through CS-UCS pairing, this experiment also tests two other hypotheses of what it is. First, Geer (1964) has suggested that it is an orienting response. If so, its size and latency should be unaffected by the changes in the ISI. Second, it may be a response due to previous conditioning which is mediated by anticipation and appears here because of the long ISI. This is similar to Steward's position (Zeaman & Smith, 1965). It was noted in the data of a previous experiment (Obrist et al., 1965) that a large deceleration appeared promptly, often on the first UCS trial before CS-UCS pairing had occurred. This suggested that deceleration was not due to CS-UCS pairing but to the instructions given *S*s before the first conditioning trial that they would see two lights as they had on adaptation trials but that one of those lights would now sometimes be followed by shock. These instructions could have created a state of anticipation following the CS, and a deceleratory response mediated by this anticipation could have developed during the long ISI. If so, deceleration should always peak near UCS onset regardless of ISI since anticipation would be greatest then. Also, it should be smaller with an .8-sec. ISI since there would be little time for anticipation and both larger and sustained with a 13-sec. ISI since the anticipatory period would be longer. These predictions are contrary to those of all other hypotheses tested here.

---

[2] The shortest interval was .8 sec. and not .5 sec., as reported earlier (Hastings & Obrist, 1965) and as specified. The shorter interval is obviously required by the experimental design, but the argument is unchanged since even .8 sec. should give more rather than less response than 13 sec.

This experiment is also a partial test of another possibility. If the deceleration is not a CR, a further question arises: why an acceleratory CR similar to the UCR fails to appear. Obrist (1965) has found a conditioned GSR and a conditioned acceleratory response that is observable under atropine but is usually masked by the dominant vagal deceleration (Samaan, 1934). Since the vagus is only relatively dominant (Cannon & Rosenblueth, 1937) and since an acceleratory UCR does appear, Obrist's findings suggest that the acceleration may be a weak CR that was inadequately developed because the ISI was so long. If so, an acceleratory CR might appear with an .8-sec. ISI. This would be further evidence that the deceleration is not a CR. It is not a test of stimulus substitution theory, however, since an acceleratory CR might still fail to appear with an .8-sec. ISI for several reasons even if conditioning were improved and the CR were acceleratory.

No previous heart-rate conditioning studies have varied the ISI systematically in this way or have used a systematic variation to test these hypotheses. Bersh, Notterman, and Schoenfeld (1953) used randomly varied intervals but did not maintain different intervals for different groups throughout conditioning. McAllister, Farber, and Taylor (1954) varied CS-UCS intervals but added a ready signal which was always given irregularly 7, 8, or 9 sec. pre-UCS regardless of whether the CS was .5 or 5 sec. pre-UCS, confounding the interval effect. Solomon (1958) used ISIs of .5, 5, and 20 sec. for different groups, but his data do not answer our questions since he averaged heart rate for pre-CS, post-CS and post-UCS periods, used no test trials, used a very intense UCS, and did not control respiration. Church and Black (1958) varied the ISI with dogs, but the measures they used would reflect only acceleration and ignore any possible deceleration. Black, Carlson, and Solomon (1962) varied the ISI with curarized dogs but report only that the speed and amount of conditioning was unchanged. They found deceleration in one dog when trace conditioning and an ISI of 10 sec. were used but evidently did not explore this further. Deane (1965) used onset-to-onset ISIs of 3, 5, and 7 sec. with rabbits and found a deceleratory CR which increased in latency as ISI increased. He attributed this to Pavlovian inhibition of delay. However, he did not attempt to test the hypotheses proposed for this research or use the extreme ISIs used here which are necessary to test them.

METHOD

Subjects.—The *S*s were 45 healthy, male undergraduates from an introductory psychology course at the University of North Carolina. Fifteen *S*s were randomly assigned to each of three experimental groups. Three other *S*s were run, but their records could not be used for technical or procedural reasons.

Apparatus.—The apparatus was identical to that used in an earlier study (Obrist et al., 1965). Heart rate, GSR, and respiration were continuously recorded by an Offner Type-R dynograph.

Procedure.—The procedure was essentially the same as that used for the group receiving low shock in an earlier study (Obrist et al., 1965). Each *S* had a rest period, respiratory training, shock level determination, adaptation to CS⁺ and CS⁻ (red and blue lights), and 16 conditioning trials with 10 test (CS⁺ only) and 10 control (CS⁻ only) trials interspersed. There were three differences from former procedures. First, three groups were used with ISIs, respectively, of .8, 7, and 13 sec. from CS onset to UCS onset. Second, no injections were given. Third, the shock level used was slightly lower. The shock determination series was stopped at "intense" rather than at "very painful" or at 1.5 ma., whichever came first. This reduced the possibility that *S* might give an involuntary gasp when the CS⁺ appeared, especially with the short .8-sec. ISI. Because of these differences, the 7-sec. group was run as an experimental control to ensure that earlier results would still be found.

RESULTS

Measurements of heart rate, respiration, and GSR were analyzed by methods which were used in previous studies (Obrist, 1963; Obrist et al., 1965; Wood & Obrist, 1964). The results in general support the fourth hypothesis: that the deceleration is not a CR developed in this experiment but is mediated by anticipation.

Heart rate.—Average values for each of 25 sec. post-CS onset were subtracted from the average for a pre-CS base period separately for test, control, and UCS trials in each group. Since the unit of measurement is milliseconds between R waves, larger values occur with a slower heart rate and deceleration gives a negative value. Figure 1 and Table 1 show average responses on test trials corrected for possible sensitization effects by subtraction of responses on control trials. The .8-sec. group shows no reliable response. The 7-sec. group shows a nonsignificant early acceleration followed by a sharp deceleration starting on Sec. 6. These results are like those of the low shock, saline group in an earlier study (Obrist et al., 1965), indicating that the slight changes made in conditions did not materially affect the heart-rate response. Test minus control-trial values are significantly different from zero on Sec. 9 through 15. The 13-sec. group shows an early acceleration reaching significance on Sec. 4, followed by a

TABLE 1

BETWEEN-GROUP *T* TESTS, T-C VALUES

	.8 Sec. vs. 7 Sec.	7 Sec. vs. 13 Sec.	.8 Sec. vs. 13 Sec.
1	0.677	−0.145	−0.876
2	0.153	−0.653	−1.149
3	0.473	0.130	−1.042
4	−0.155	−0.978	−0.964
5	0.769	−0.104	−1.249
6	−0.149	−0.691	−0.814
7	−1.952	−0.981	1.364
8	−2.690*	−0.599	3.091**
9	−2.535*	−1.499	1.898
10	−3.403**	−2.187*	1.801
11	−4.565**	−2.129*	2.785**
12	−4.144**	−1.136	3.578**
13	−3.135**	−1.285	2.183*
14	−2.600*	−1.275	1.668
15	−2.227*	−0.398	2.544*
16	−1.415	0.582	2.586*
17	−1.178	1.070	3.103**
18	−1.049	1.318	2.880**
19	−0.906	1.324	2.258*
20	−1.648	1.099	2.731*
21	−1.696	0.752	2.593*
22	−1.782	0.659	2.578*
23	−0.737	0.567	1.313
24	−1.358	0.000	1.311
25	−0.573	0.490	1.027

* $p < .05$ (2.05).
** $p < .01$ (2.76).

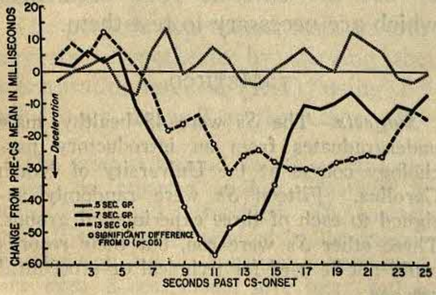


FIG. 1. Heart-rate responses on test minus control trials.

more gradual and sustained deceleration starting on Sec. 5 and peaking on both Sec. 12 and 18. Test minus control-values are significantly different from zero on Sec. 8, 9, and 11 through 22.

Differences between groups are also reliable. The response of the .8-sec. group differs significantly from the deceleration of the 7 sec. one on Sec. 8 through 15 and from the deceleration of the 13 sec. one on Sec. 8, 11 through 13, and 15 through 22. Differences between the 7- and 13-sec. groups are significant on Sec. 10 and 11.

Although this was the major analysis originally planned for this study, recent evidence indicates that uncorrected test-trial values may be a more appropriate measure of conditioning. A study of heart-rate changes following the CS in a procedure similar to this except that neither CS was paired with the UCS (Obrist & Wood, 1966) indicates that changes on CS or control trials may be due to generalization rather than to sensitization. A comparison of responses on test and control trials in the present experiment supports this view. As Fig. 2 shows, test and control trial responses are roughly parallel for the three groups. There is more deceleration on test trials in the 7- and 13-sec. groups, but the latency and duration of the deceleration also increases with the ISI on control trials. In the .8-sec. group, heart rate returns to the base level on Sec. 5 on control trials and Sec. 6 on test trials. With a 7-sec. ISI, it returns by Sec. 13 on control trials and 16 on test trials. And with a 13-sec. ISI, it returns on Sec. 20 on control trials and has not returned by Sec. 25 on test trials. This is clear evidence of generalization since sensitization should cause a CS-bound response unaffected by the ISI.

Test-trial responses may, therefore, reflect the effect of conditioning better than test minus control-trial values. The .8-sec. group shows a new biphasic response on test trials opposite in direction to that shown by 7-sec. groups without respiratory control: a small, short deceleration significant on Sec. 3, followed by a larger acceleration significant on Sec. 8 through 25. The 7- and 13-sec. groups show a shift in peak response time similar to that found on corrected test trials. The 7-sec. group shows a sharp deceleration significant on Sec. 7 through 14 and peaking on Sec. 9 through 11, near UCS onset. The 13-sec. group shows a more gradual and shallower but much more sustained deceleration significant from Sec. 7 through Sec. 22 and peaking on Sec. 15 and 16, also near UCS onset. Inspection of the records suggests, however, that this slightly shallower, sustained deceleration is due to the averaging of somewhat shorter large responses peaking at different

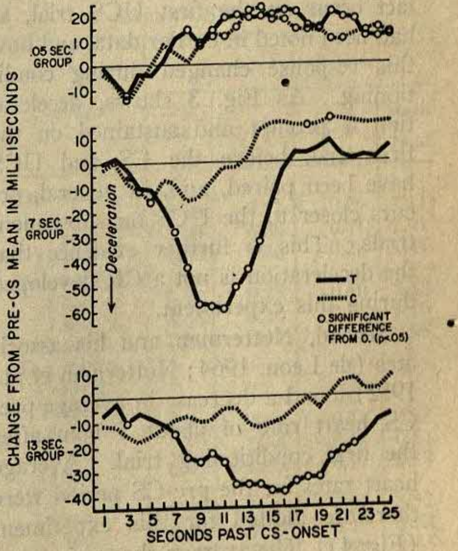


FIG. 2. Heart-rate responses on test and control trials.

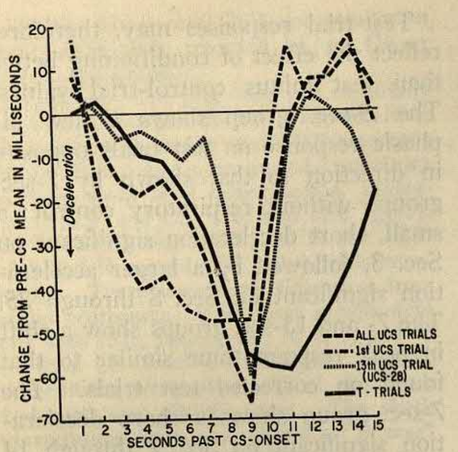


FIG. 3. Heart-rate responses on the first and thirteenth UCS trials and on all test and UCS trials (7-sec. group).

times rather than a diminution of response size.

Two additional analyses were carried out for the 7-sec. group only to see whether these results are similar in two other effects to the results of earlier studies. First, selected UCS trials were analyzed separately to determine whether deceleration did in fact occur on the first UCS trial, as had been noted in earlier data, and how this response changed during conditioning. As Fig. 3 shows, deceleration is prompt and sustained on the first trial, before the CS and UCS have been paired, and, in general, occurs closer to the UCS on subsequent trials. This is further evidence that the deceleration is not a CR developed during this experiment.

Second, Notterman and his associates (de Leon, 1964; Notterman et al., 1952) found a decrease in average pre-CS heart rate of about 5 bpm after the first conditioning trial. Average heart rates for the pre-CS period were therefore plotted for this experiment (Fig. 4). The average change in heart rate was not only sudden (63 msec. between the first two trials) but more sustained than the effect Notterman noted, which lasted only a few trials. The average change was 49 msec. (roughly 5 bpm) on 25 trials (only 6 UCS trials were sampled) with $p < .001$ ($t = 15.99$). One S had an average change of roughly 20 bpm. This pre-CS change cannot be due to preshock acceleration since the pre-CS level on UCS-1 is only 10 msec. faster than the pre-CS level on the last adaptation trial. It is impressive because the deceleration previously discussed occurred in addition to this overall change of base-level heart rate. It is also important because it is further evidence that situational factors may change heart rate. This pre-CS change can hardly be due to simple CS-UCS pairing.

Respiration.—Since the measure of respiration used here could not be calibrated, respiration was quantified as the ratio of the largest inspiration in certain post-CS onset intervals to the mean inspiration in a 15-sec. pre-CS base period when respiration was also controlled. The post-CS intervals were Sec. 1–6 for all groups, Sec. 7–15 for the .8- and 7-sec. groups, and Sec. 7–12 and 13–21 for the 13-sec. group. This gave post-CS onset and post-

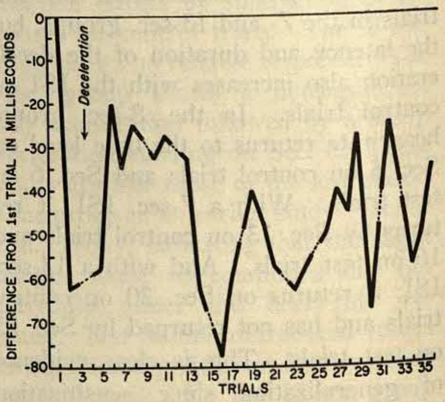


FIG. 4. Mean differences of basal pre-CS heart-rate values from the pre-CS value on the first conditioning trial.

TABLE 2

Post-CS Respiratory Changes

Trial Type	ISI Group						
	0.8 Sec.		7 Sec.		13 Sec.		
	F6[a]	L9[b]	F6[a]	L9[b]	F6[a]	M6[c]	L9[d]
T	1.10*	1.11**	1.13**	1.07	1.07	0.99	1.00
C	1.10**	1.10**	1.04	1.08*	1.00	0.98	1.04
T-C	0.00	0.01	0.09*	−0.01	0.07	0.01	−0.04
UCS	1.34**	1.28**	1.07	1.38**	1.06	1.01	1.32*

[a] Seconds 1–6 post CS-onset.
[b] Seconds 7–15 post CS-onset.
[c] Seconds 7–12 post CS-onset.
[d] Seconds 13–21 post CS-onset.
* $p < .05$ by t test.
** $p < .01$ by t test.

UCS onset values for all groups. As Table 2 shows, the results do not indicate that the heart-rate changes found in the three groups are consistently related to respiratory changes. For example, on test trials the values are almost the same for both periods for the .8- and 7-sec. groups yet the heart-rate changes are very different. If small differences are important, then respiratory changes on T and UCS trials for the 7-sec. group should not cause almost identical deceleration (Fig. 3).

To check this further, two other analyses were carried out on the 7-sec. group only. First, the ratio of the largest inspiration during the pre-CS period to the mean for that period was determined and was found to be 1.23. This is a value for the preperiod comparable to the values above for post-periods and since it is actually larger than they are, except during shock, it is further evidence that maximum-respiratory amplitude did not increase following CS onset. However, since this value was larger, it seemed possible that the post-CS deceleration of heart rate found in this group might be due to an inhibition of breathing after the CS appeared. The ratios of post-CS means for Sec. 1–6 and 7–15 to the pre-CS mean were therefore determined for all trial types and were found to be approximately 1.00 except during shock. In other words, there was no difference in average breathing depth between pre- and post-CS periods. These findings, considered together, indicate that breathing was better controlled and more uniform after CS-onset and that the deceleration found in this group cannot be due to either abnormally deep breaths or unusually shallow breathing.

GSR.—The GSR, unlike heart rate, showed the large response with a .8-sec. ISI and progressively smaller responses with longer ones predicted by stimulus-substitution theory. A response was defined as a decrease in resistance of at least 600 ohms. Table 3 gives the average response per trial for the intervals 1–6, 7–12, and 13–19 sec. post-CS onset. The .8-sec. group has a much larger response than either of the others for the first period on both test and control trials. The difference between the .8- and 13-sec. groups is

TABLE 3

Mean GSR Changes

Group	Seconds Post-CS Onset			
	1–6	7–12	13–19	Mean of Total
Test Trials				
.8 sec.	8.88	0.17	0.29	3.11
7 sec	5.05	0.56	0.11	1.91
13 sec.	2.82	0.36	0.28	1.11
Control Trials				
.8 sec.	6.94	0.25	0.40	2.53
7 sec.	2.86	0.20	0.23	1.10
13 sec.	2.10	0.41	0.45	0.98
Test minus Control Trials				
.8 sec.	1.94	−0.08	−0.11	0.58
7 sec.	2.19	0.36	−0.12	0.81
13 sec.	0.72	−0.05	−0.17	0.13
UCS Trials				
.8 sec.	18.97	0.40	0.21	6.52
7 sec.	8.78	7.99	0.17	5.63
13 sec.	3.87	0.34	11.12[a]	4.65[a]

[a] Results of only 7 *S*s because of technical difficulties.

significant at the .001 level on both test and control trials and the difference between the .8- and 7-sec. groups is significant at the .05 level on control trials and approaches significance ($p < .10$) on test trials (two-tailed Mann-Whitney *U* test).

Also, first responses for the .8-sec. group had an average latency of 2.35 sec. so that they occurred at the same time as the early heart-rate deceleration found in this group. This evidence of combined sympathetic and parasympathetic activation will be discussed later.

DISCUSSION

These results clearly support the hypothesis that the deceleration found here is a previously conditioned response mediated by anticipation and they fail to support any of the others. First, they indicate that it is not a CR developed during this experiment. The deceleration is much smaller with a usually more optimal .8-sec. interval (Kimble, 1961) and larger, peaking later, with a long, 13 sec. one which should produce little conditioning. Also, the response does not appear after repeated CS-UCS pairings but actually occurs on the first UCS trial, before the CS and UCS have been paired. Second, these data indicate that it is not an orienting response reactivated by shock (Geer, 1964) since an orienting response should be unaffected in size and latency by changes in the ISI and this response varies in both. It is especially difficult to understand why an orienting response would become appreciably smaller with a short ISI and larger, peaking later, with a long one. Third, the results do not support the possibility that an acceleratory CR would appear with a more optimal .8-sec. ISI. However, they do not eliminate the possibility that the acceleration previously found under atropine (Obrist et al., 1965) was augmented by the short interval although it still failed to appear. This suggests a further test of the .8-sec. group under atropine. Fourth, the data indicate that deceleration is unrelated to either an increase or decrease in respiratory amplitude following the CS.

The hypothesis of mediation by anticipation, on the other hand, can account for all of the results. First, the major finding, a smaller heart-rate deceleration with an .8-sec. ISI and a larger, later one with a 13-sec. one, was actually predicted by this hypothesis since there would be little time for anticipation with a short interval and there would be more anticipation, peaking near UCS-onset, with a longer one. No other hypothesis predicted this result. Second, the large, significant deceleration on the first UCS trial, before CS-USC pairing, which is unexplained by the other hypotheses, can be explained as due to anticipation engendered by instructions that *S*s would now sometimes receive shock. The shift in latency of this response can be explained as due to *S*s' learning the length of the ISI. On the first trial, deceleration begins soon after CS onset because *S*s do not know when the shock will occur. Later it begins nearer UCS onset because they have learned that it comes later. This also explains the tendency for *S*s to peak at the same time with a 7-sec. ISI but peak at slightly different times with a 13 sec. one: they may not judge the length of the longer ISI as well. Third, the parallel response on T and C trials probably reflects generalization, which could occur as easily with this type of response as with a typical CR. Fourth, the lower pre-CR averages after the first UCS trial can be explained as reflecting anticipation of the next CS. Anticipation would be greater after the first shock since the CS is a more meaningful stimulus. Finally, the conditioned GSR, which always occurs near CS onset and is larger with an .8-sec. ISI and smaller with a 13 sec. one, suggests strongly that the deceleration found here may occur in addition to and may mask a sympathetic acceleratory CR that would reflect ISI changes as the GSR does, being larger with an .8-sec. interval but still not strong enough to override the vagal deceleration. This possibility is not antagonistic to the fourth hypothesis but is further evidence that the deceleration is not the CR developed here.

This hypothesis can also explain some of the more puzzling results of earlier experiments. Sherrington's dog (Sherrington, 1900) decelerated when it heard a vibrator that was always turned on shortly before the dog was shocked. The deceleration could therefore be due to anticipation of being shocked. Kling's cats (Kling, 1933) may have shown anticipatory fear of what would happen next after they were confined. Some other experiments also may contain an overlooked element of anticipation which could have caused deceleration when excessive movement did not cause an overriding acceleration. For example, Black (Black et al., 1962) reports both acceleration and deceleration in curarized dogs. The deceleration appeared more frequently with trace conditioning, in which there is a period without a stimulus when anticipation might develop, than with delay conditioning, in which the CS may take on the noxious properties of the UCS so that there is immediate response to it rather than pure anticipation of the UCS. In both Notterman's experiments (Notterman et al., 1952) and Zeaman's early work paralleling them (Zeaman, Deane, & Wegner, 1954), *S*s were instructed that they would receive shock "a little later in the experiment." They therefore would not anticipate shock until after the first shock occurred but, as Notterman's results clearly show, a large deceleratory response then appeared in his experiment beginning with the next trial. In fact, the deceleration on the second trial was the largest Notterman obtained during conditioning. Zeaman also found unusually fast "conditioning" with maximum deceleration at about the fifth CS-UCS pairing. Both results fit the mediation by anticipation hypothesis better than a stimulus substitution one. Finally, this hypothesis also explains an unnoticed effect in Zeaman's respiratory studies (Zeaman & Smith, 1965): All *S*s decelerated near the end of sustained inspiration or expiration if they knew how long the period would last, but they did not decelerate during the preconditioning-control periods when they could not anticipate the end of the period. This effect was especially strong during sustained exhalation, when anticipation of the end of the period would be particularly great, and it suggests that failure to find a large deceleration near the UCS may have been partially due to preoccupation of *S*s with the respiratory control.

The hypothesis of mediation by anticipation leaves us with a further question, however: why anticipation of shock should cause deceleration. Although these results explain the seeming conflict with stimulus-substitution theory, they still fail to fit Cannon's theory. The deceleration found in other recent experiments also fails to fit his theory. Deceleration may occur in anticipation of a nonnoxious stimulus: Steward's pleasant sound (Zeaman & Smith, 1965), or pictures of nudes and exercise (Wood & Obrist, 1965). Also, Lacey (Lacey, Kagan, Lacey, & Moss, 1963) and, later, Obrist (1963) have found deceleration during nonnoxious stimuli. Lacey has suggested that this deceleration is related to a wish to take in the environment and that the acceleration found during noxious stimuli or mental concentration is related to a wish to reject the environment. This excellent suggestion does not explain, however, why deceleration occurs in anticipation of a noxious stimulus which *S* would wish to reject. Also, if the wish to reject the environment by itself causes acceleration, then acceleration should occur during rest when the wish to reject the environment is also strong. It is therefore necessary to modify Lacey's theory as well as Cannon's to fit all data.

Another major flaw in Cannon's theory is that it does not fit readily observable animal behavior. The senior author has watched wild animals for many years and has found that ongoing activity, when the parasympathetic system is presumably dominant, is coupled with extreme peripheral alertness. Anyone can varify this observation by watching birds at a feeder. It is also evident that this must be so.

Even when there is no immediate danger, all wild animals must maintain constant anticipatory alertness for stimuli anywhere in the environment warning of a predator or a rival for food, mate, or territory as well as stimuli telling about food or a mate. Although the mechanism linking the parasympathetic dominance of ongoing activity with this alertness undoubtedly developed by chance (this is not a teleological argument), it was maintained because it gave its possessors a distinct evolutionary advantage. Only civilized man and the animals he protects can afford the inertness we commonly associate with parasympathetic dominance.

Observation also indicates that this diffuse alertness for potential stimuli changes during sympathetic arousal to concentration on relevant present stimuli and relative inattention to all other present or potential stimuli. Fighting or fleeing animals crash into obstacles as if they did not see them. Intense pain masks a weaker one as the dolorometer shows. When a mouse is introduced into a cat's cage responses to clicks in the cat's cochlear nucleus diminish (Hernandez-Peon, Scherrer, & Jouvet, 1956). This mechanism would also give an evolutionary advantage: in an emergency, potential dangers matter less than present ones.

These observations as well as the data of recent experiments suggest a modification of Cannon's theory. It is proposed that, instead of one passive and one active system with an organismic continuum from inertness to alert arousal as dominance shifts, we have two active systems, each arousing the organism in a specific way. In this view, inertness would occur when neither system was activated and various patterns of arousal occur depending on the degree and type of activation of each system.

Lacey's theory and data fit this view with little modification. His wish to take in the environment easily stretches to include the need to be alert for potential stimuli, pleasant or unpleasant, which is the basis of anticipatory deceleration. Actually, there is an element of either anticipation or peripheral scanning in the stimuli Lacey and Obrist used. This would explain why there is more deceleration during a flashing light (photic stimulation), a variable white noise, or a task of looking for differences between two pictures than during a simple light or tone. Also, mental concentration is an extension of the focal concentration of emergency arousal and it is possible that this, as well as the wish to reject the environment, causes the acceleration found during problem solving. This would explain why Billings and Shepard (1910) found acceleration during reading German at a distance, a task of taking in the environment but also one of concentration.

This modification of Cannon's theory would, therefore, fit all present data with a minimal change in an old and useful theory. It would also elucidate several problems in psychosomatic medicine. Needless to say, however, it also requires repeated rigorous testing.

REFERENCES

Bekhterev, V. R. Emotions as somatomimetic reflexes. In M. L. Reymert (Ed.), *Feelings and emotions.* Worchester, Mass.: Clark Univer. Press, 1928.

Beier, C. D. Conditioned cardiovascular responses and suggestions for treatment of cardiac responses. *J. exp. Psychol.,* 1940, 26, 311–321.

Bersh, P. J., Notterman, J. M., & Schoenfeld, W. N. The effect of randomly varying the interval between conditioned and unconditioned stimuli upon the production of experimental anxiety. *Proc. Nat. Acad. Sci.,* 1953, 39, 563–570.

Billings, M. L., & Shepard, J. F. The change of heart rate with attention. *Psychol. Rev.,* 1910, 17, 217–227.

Black, A. H. Cardiac conditioning in curarized dogs: The relationship between heart rate and skeletal behaviour. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965. Pp. 20–47.

Black, A. H., Carlson, N. J., & Solomon, R. L. Exploratory studies of the conditioning of autonomic responses in curarized dogs. *Psychol. Monogr.,* 1962, 76(29, Whole No. 548).

Cannon, W. B., & Rosenblueth, A. *Autonomic neuro-effector systems.* New York: MacMillan, 1937.

CHURCH, R. M., & BLACK, A. H. Latency of the conditioned heart rate as a function of the CS-UCS interval. *J. comp. physiol. Psychol.,* 1958, **51**, 478–487.

DEANE, G. E. Cardiac conditioning in the albino rabbit using 3 CS-UCS intervals. *Psychon. Sci.,* 1965, 3, 119–120.

DE LEON, G. Conditioning the human heart rate with noise as the unconditioned stimulus. *J. exp. Psychol.,* 1964, **68**, 518–520.

GANTT, W. H. *Experimental bases for neurotic behavior.* New York: Harper, 1944.

GANTT, W. H. Postscript to experimental induction of psychoneuroses by conditioned reflex with stress. In *The biology of mental health and disease: The 27th annual conference of the Millbank Memorial Fund.* New York: Hoeber, 1952. Pp. 508–514.

GEER, J. H. Measurement of the conditioned cardiac response. *J. comp. physiol. Psychol.,* 1964, **57**, 426–433.

HASTINGS, S. E., & OBRIST, P. A. Heart rate conditioning in humans: Effects of varying the inter-stimulus (CS-UCS) interval. Paper read at Society for Psychophysiological Research, Houston, October 1965.

HERNANDEZ-PEON, R., SCHERRER, H., & JOUVET, M. Modification of electric activity in cochlear nucleus during attention in unanesthetized cats. *Science,* 1956, **123**, 331–332.

KIMBLE, G. A. *Hilgard & Marquis' Conditioning and learning.* (Rev. ed.) New York: Appleton-Century-Crofts, 1961.

KLING, C. The role of the parasympathetics in emotions. *Psychol. Rev.,* 1933, **40**, 368–380.

KOSUPKIN, J. M., & OLMSTEAD, J. M. D. Slowing of the heart as a conditioned reflex in the rabbit. *Amer. J. Physiol.,* 1943, **139**, 550–552.

KOTHAREVSKY, L. I. Cardio-vascular conditioned reflex to direct and to verbal stimuli. *Fiziol. Zh. SSSR,* 1936, **20**, 228–242.

LACEY, J. I., KAGAN, J., LACEY, B. C., & MOSS, H. A. Situational determinants and behavioral correlates of autonomic response patterns. In P. J. Knapp (Ed.), *Expression of the emotions in man.* New York: International Univer. Press, 1963. Pp. 161–196.

MCALLISTER, W. R., FARBER, I. E., & TAYLOR, J. E. Conditioned heart rate as a function of anxiety and CS-UCS interval. Technical Report No. 1, 1954, Contract N9 onr-93802, Project NR 154-107, Office of Naval Research.

NOTTERMAN, J. M., SCHOENFELD, W. N., & BERSH, P. J. Conditioned heart rate responses in human beings during experimental anxiety. *J. comp. physiol. Psychol.,* 1952, **45**, 1–8.

OBRIST, P. A. Cardiovascular differentiation of sensory stimuli. *Psychosom. Med.,* 1963, **25**, 450–458.

OBRIST, P. A., & WOOD, D. M. The sensitized heart rate response in humans as evaluated by simple procedures. *J. exp. Psychol.,* 1966, **72**, 779–781.

OBRIST, P. A., WOOD, D. M., & PEREZ-REYES, M. Heart rate during conditioning in humans: Effects of UCS intensity, vagal blockade, and adrenergic block of vasomotor activity. *J. exp. Psychol.,* 1965, **70**, 32–42.

PETERS, J. F., & GANTT, W. H. Conditioning of human heart rate to graded degrees of muscular tension. *Fed. Proc.,* 1951, **10**, 104.

REESE, W. G., & DYKMAN, R. A. Conditioned cardiovascular reflexes in dogs and man. *Physiol. Rev.,* 1960, **40**, 250–265.

ROBINSON, J., & GANTT, W. H. The orienting reflex (questioning reaction): Cardiac, respiratory, salivary and motor components. *Bull. Johns Hopkins Hosp.,* 1947, **80**, 231–253.

SAMAAN, A. The antagonistic cardiac nerves and heart rate. *J. Physiol.,* 1934–35, **83**, 332–340.

SCHOENFELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 70–99.

SHERRINGTON, D. C. Experiments on the value of vascular and visceral factors for the genesis of emotion. *Proc. Roy. Soc. Lond.,* 1900, **66**, 390–404.

SMITH, R. W., & ZEAMAN, D. Heart rate during anxiety. Progress Report No. 1, March 1964, Research Grant MH-6765, National Institute of Health.

SOLOMON, R. L. Some autonomic correlates of human traumatic avoidance learning. Progress Report, Project M-1246, June 1958, United States Public Health Service.

SSOROKHTIN, G. N., TURGEL, K. U., & MINUT-SSOROKHTIN, O. P. The conditioned vagotropic reflex of the heart. *Fiziol. Zh. SSSR,* 1934, **17**, 207–216.

WOOD, D. M., & OBRIST, P. A. Effects of controlled and uncontrolled respiration on the conditioned heart-rate response in humans. *J. exp. Psychol.*, 1964, **68**, 221–229.

Wood, D. M., & Obrist, P. A. Heart rate during conditioning in humans: Effects of unconditioned stimuli with minimal and maximal sensory input. Paper read at Society for Psychophysiological Research, Houston, September 1965.

Zeaman, D., & Smith, R. W. Review and analysis of some recent findings in human cardiac conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1964. Pp. 378–418.

(Received July 22, 1966)

(Continued from page 430)

Subjective Probabilities Inferred from Estimates and Bets: Lee Roy Beach* and Lawrence D. Phillips: Department of Psychology, Denny Hall, University of Washington, Seattle, Washington 98105.

Delay of Positive Reinforcement in Instrumental Eyelid Conditioning: Louise E. Cerekwicki* and David A. Grant: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.

Unlearning of List-1 Right Items in Verbal-Discrimination Transfer: Donald H. Kausler,* Frank E. Rulderson, and A. John Eschenbrenner, Jr.: Department of Psychology, Saint Louis University, 221 North Grand Boulevard, Saint Louis, Missouri 63103.

Effects of Formal Interitem Similarity and Length of Retention Interval on Proactive Inhibition of Short-Term Memory: John H. Wright*: Department of Psychology and Sociology, Virginia Polytechnic Institute, Blacksburg, Virginia 24061.

Resistance to Extinction When a Single Nonreinforced Sequence is Followed by Immediate Reinforcement or Delayed Reinforcement: Glen D. Jensen* and Winfred F. Hill: Department of Psychology, DePaul University, 1219 West Fullerton Avenue, Chicago, Illinois 60614.

Temporal Integration: Modification of the Incentive Value of a Food Reward by Early Experience with Deprivation: K. Edward Renner*: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Effects of Postresponse Stimulus Duration upon Short-Term Memory: Norman R. Ellis* and Terry R. Anders: Department of Psychology, University of Alabama, University, Alabama 35486.

Instructions about Randomness and Run-Dependency in Two-Choice Learning: Lee Roy Beach* and Richard G. Swensson: Department of Psychology, University of Washington, Seattle, Washington 98105.

Surrogate Processes in the Short-Term Retention of Connected Discourse: Kenneth F. Pompi and Roy Lachman*: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

Retroactive Inhibition as a Function of Learning Method: Thomas J. Shuell and Geoffrey Keppel*: Department of Psychology, University of California, Berkeley, California 94720.

Performance Changes in Escape Conditioning Following Shifts in the Magnitude of Reinforcement: Paul J. Woods*: Department of Psychology, Hollins College, Hollins, Virginia 24020.

Effects of Discrimination Training on Stimulus Generalization for Human Subjects: Theodore J. Doll and David R. Thomas*: Department of Psychology, 215 Kent Hall, Kent State University, Kent, Ohio 44240.

Stimulus Learning and Recognition in Paired-Associate Learning: Harley A. Bernbach: Ventura Hall, Stanford University, Stanford, California 94305.

Hypothesis Sampling in Concept Identification: James R. Erickson*: Department of Psychology, Human Performance Center, 404-B Seventeenth Avenue, Ohio State University, Columbus, Ohio 43210.

Effect of Supplemental Visual Cues on Rotary Pursuit: Norman B. Gordon* and Merrill J. Gottlieb: Department of Psychology, Yeshiva University, 55 Fifth Avenue, New York, New York 10003.

(Continued on page 451)

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 443–451

DIFFERENTIAL RETENTION OF INDIVIDUAL PAIRED ASSOCIATES WITHIN AN R T T "LEARNING" TRIAL[1]

TANNIS Y. ARBUCKLE

University of Toronto[2]

Single presentations of lists of paired associates followed by 1 or 2 recall tests were used to evaluate the hypothesis that performance on an R T T trial is determined solely by how well items are retained during the intervals separating presentation and Test 1 (T_1), and T_1 and Test 2 (T_2). 6 experiments, employing various combinations of list length and presentation rate, were conducted in which lengths of the 2 retention intervals were varied by manipulating each item's serial position in the T_1 and T_2 test sequences relative to its position in the study sequence. Relative frequency of correct response on T_1 and of forgetting and recovery on T_2 was found to vary systematically with position in the study and recall sequences. Implications of these results for the original hypothesis and for the incremental and all-or-none models of learning are discussed.

A small amount of verbal material can virtually always be recalled if tested immediately after presentation (Melton, 1963; Murdock, 1961; Peterson, 1963; Waugh & Norman, 1965). As Tulving (1964) has pointed out, this finding suggests that learning is neither all or none nor incremental, but rather, simply *all*. If one takes the view that each item on a list is learned when presented, then the problem of accounting for observed variability in the recall of different items following a single presentation of the list becomes one of specifying the factors which cause certain items to be forgotten. When lists of items are being learned, there is almost always an interval between when an item is presented and when it is tested for recall, an interval in which rapid forgetting could occur. In an R T T "learning" trial (Estes, 1960) there are two such retention intervals, the first (I_1) comes between presentation and the first recall test (T_1), and the second (I_2) between T_1 and the second recall test (T_2). For any item, I_1 may be filled with the presentation (input) and attempted recall (output) of other items, while I_2 may be filled only with output of other items.

Tulving and Arbuckle (1963), using lists of 10 paired-associate (PA) items presented at a 3-sec. rate, showed that the input and output of other list items could result in failure to recall any given item on T_1. The present series of experiments adapted the method of Tulving and Arbuckle to the R T T trial. The use of the R T T paradigm permitted evaluation of the effects of intratrial activity on the forgetting and recovery of individual items from T_1 to T_2. It was expected that such a detailed analysis of the "fate" of individual items over successive recall tests would provide a means to assess the various theoretical positions, "all," "all or none," and "incremental" that could be taken with respect to data from R T T experiments.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Toronto. The investigation was supported in part by the National Science Foundation under Grant No. GB-810 to G. Mandler and E. Tulving, and in part by predoctoral awards made to the author by the Canada Council and by the National Research Council of Canada. The author is greatly indebted to E. Tulving for the aid and encouragement given during the course of this work.

[2] Now at Sir George Williams University, Montreal, Canada.

METHOD

Six experiments were conducted, each being a partial replication of the others. Experiments I, II, and III used list lengths of 12, 9, and 6 items, respectively, and held presentation rate constant at 3 sec. per item. Experiments IV, V, and VI used presentation rates of 2, 3, and 4 sec. per item, respectively, and held list length constant at 6 items. Thus, the conditions of Exp. V exactly replicated those of Exp. III.

Materials

Twelve different lists of 12 A-B pairs were constructed, each list having the Numbers 4–15 as A members, and common nouns belonging to the same conceptual category as B members. For each category the 12 most common words of 3–6 letters, as tabulated by Cohen, Bousfield, and Whitmarsh (1957), were paired randomly with the Numbers 4–15 to yield the 12 lists.

Experiment I used all 12 lists of 12 pairs. Mean recall per category was calculated and the categories were ordered in descending order of these means as follows: (*a*) Clothing, (*b*) Animals, (*c*) Vehicles, (*d*) Professions, (*e*) Metals, (*f*) Trees, (*g*) Fruits, (*h*) Fabrics, (*i*) Birds, (*j*) Vegetables, (*k*) Sports, and (*l*) Flowers.

Experiment II used 9 lists of 9 pairs—Categories *b–j* inclusive and pairs within categories whose A members went from 5 to 13 inclusive.

Experiments III–VI all used 6 lists of 6 pairs—Categories *d–i* inclusive, and pairs within categories whose A members went from 7 to 12 inclusive.

Twenty-four pairs of uncommon nouns and adjectives were prepared as practice items.

All A-B and A items were printed on 2 × 2 negative slides. Letter size was such that when items were projected by a Carousel projector on a flashed-opal viewing screen, the letters viewed by *S* were about 1 in. high and well spaced.

Design

The recall method of PA learning (Battig & Brackett, 1962) was used to separate the input and output phases of the R T T trial.

Input sequences.—In the input phase, pairs were presented in ascending order of A members, the "ordered" input condition of Tulving and Arbuckle (1963).

Output sequences on T_1.—Output sequence refers to the order in which items, specified by their A members, are tested for recall. The number of different T_1 output sequences constructed for each experiment was *L*, where *L* equals the length of the lists being used. The method of constructing these sequences followed that outlined by Tulving and Arbuckle (1963, pp. 322–323) except that (*a*) output sequences in Exp. II were generated by a 9 × 9 Latin square and (*b*) output sequences in Exp. IV–VI were based on random sequence of A members, and not on the input sequence as had been done in the Tulving and Arbuckle study and in Exp. I–III. Despite these exceptions, the essential requirement of the Tulving and Arbuckle method was met in that for each *S*, items presented in each input position were tested once on T_1 in each of the possible T_1 output positions. Under these conditions, each *S* contributed one datum, a correct response, an intrusion, or an omission, to each cell of a T_1 serial position (SP) matrix, where the *L* rows represented successive positions in the input sequence, the *L* columns, successive positions in the T_1 output sequence, and the L^2 cells, all possible combinations of input position and T_1 output position. Since each possible combination produced an I_1 which differed from all others in terms of number and combination of inputs and outputs it contained, the result was that recall of items was tested once for each *S* following each kind of I_1.

Output sequences on T_2.—The addition of T_2, through use of the R T T paradigm, created L^3 different possible combinations of input position and T_1 and T_2 output position. To avoid prolonging the experimental sessions unduly, only one-third of the items, representing $L^2/3$ of these different possible combinations, was actually retested on T_2. The items sampled were those appearing in every third output position in the T_1 output sequence, commencing with the second T_1 output position.[3] These items were retested on T_2 in reverse order to the order of testing on T_1. This method of sampling meant that, over the *L* lists within an experiment, each

---

[3] Postexperimental examination of the T_1 data indicated that the particular items sampled on T_2 did not differ in their T_1 recall from the items not so sampled.

S contributed one datum to each cell of a T_2 SP matrix, where the *L* rows represented successive positions in the input sequence and the *L*/3 columns, the particular combinations of T_1 and T_2 output positions that were used.

From *S*'s viewpoint, the T_1 and T_2 sequences constituted a single output sequence, as there was no pause between the two phases. A typical example of such an output sequence, taken from Exp. III was: 9-, 12-, 8-, 11-, 7-, 10-, 7-, 12-.

Subjects

The *S*s were 300 University of Toronto undergraduates who participated as part of their course requirements. Fifty *S*s were assigned nonsystematically to each experiment, with the restriction that in each there be 23 males and 27 females, this being the approximate sex ratio in *S* pool. The median age of *S*s was 19.0 yr.

Procedure

All *S*s were tested individually. They were instructed to read aloud all A-B and A items as they were shown. Recall was oral, and all responses were recorded by *E* provided they were given before the next A item appeared. Late responses were recorded as omissions.

Two preliminary trials, each using *L* different practice items, were given to ensure comprehension of the instructions and to reduce initial warm-up and learning-to-learn effects. The *S*s were then told the general nature of the experimental materials and were warned that some items might be tested more than once. Prior to presentation of each list, *S* was told the particular category represented by its B members. After the recall of each list, *S* was told how many correct responses he had made on T_1 and T_2 combined.

Slides were presented throughout each experiment at the rate appropriate to the experimental conditions. The stated rate included an interslide interval of .75 sec. Blank slides, handled at the same rate, were inserted at the beginning and end of the input sequence to alert *S* to the beginning of the input and output phases, respectively.

Order of presentation of experimental lists was varied systematically across *S*s so that, within each experiment, each list was presented approximately equally often in each position in the series of lists. For each *S*, the *L* different lists were associated with the *L* different output sequences according to a table of random numbers.

RESULTS

T_1 *recall.*—Relative frequency of correct recall, averaged across items and *S*s, ranged from .30 in Exp. I to .50 in Exp. VI. In general, correct recall increased with shorter list lengths and slower presentation rates. This increase was accompanied by a corresponding decrease in response omissions. Relative frequency of intrusions varied only slightly, from .22 to .27 across the six experiments. Of these intrusions, over 99% came from within the list.

The effects of input and T_1 output position on correct recall were evaluated for each experiment by analysis of variance on the arc sin transformations of the relative frequencies of recall in the L^2 cells of the T_1 SP matrix. The analysis showed that, in each of the experiments, the effects of input position, of T_1 output position, and of their interaction were all significant ($p < .01$ in all cases).

The nature of these effects is illustrated in Fig. 1 for selected combinations of input and T_1 output positions. [Complete tabulations of the data in both the T_1 and T_2 SP matrices are available in Arbuckle (1964).] In all experiments the first and last inputs were better recalled than middle ones. The consistently high recall of first inputs agrees with Tulving and Arbuckle's (1963) finding of a pronounced primacy effect with "ordered" lists. Output position affected recall of last inputs, but not that of first or middle inputs.

The finding that input position and T_1 output position systematically affect T_1 recall still leaves open the question of whether, for a given length of I_1, interpolated outputs affect recall dif-

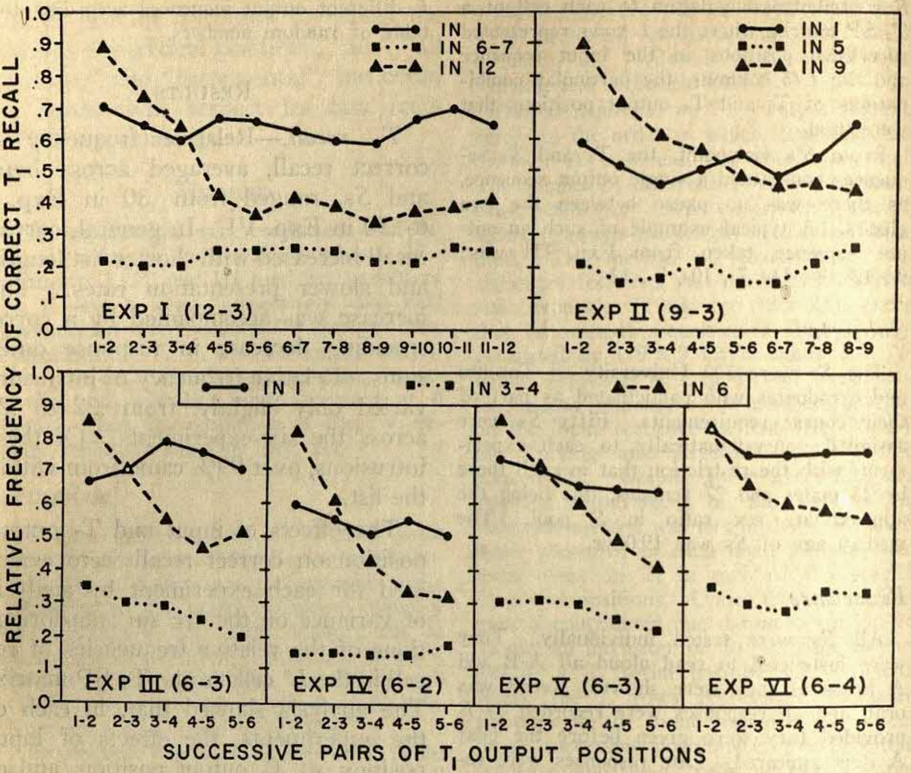


FIG. 1. Relative frequency of correct recall on T_1 averaged over successive pairs of T_1 output positions with input position as parameter for Exp. I–VI.

ferently from interpolated inputs. This question may be answered by comparing, for each experiment, the recall of the last item in the input sequence where I_1 can contain only interpolated outputs, with the recall of the first item in the output sequence where I_1 can contain only interpolated inputs, for each length of I_1 from 0 to $L - 3$. (The first two input positions and hence the last two output positions have been excluded from this comparison because of the pronounced primacy effect.) In all experiments, as Fig. 2 shows, with length of $I_1 = 0$, relative frequency of correct recall was near 1.0, and with each length of $I_1 > 0$, relative frequency of correct recall was higher following interpolated outputs than following interpolated inputs.

T_2 recall.—Table 1 shows T_2 responses tabulated in relation to the preceding T_1 response. In each experiment, intertest forgetting (CN) was about equal to intertest recovery (NC), with the relative frequencies of both varying little across experiments. Even though both kinds of response shift were much less frequent than were successive correct (CC) or successive incorrect (NN) responses, they were both observed to occur relatively more frequently under some combinations of input position and length of I_2 than under others. It is with these systematic variations in CN and NC responses that all further analysis of the results is concerned.

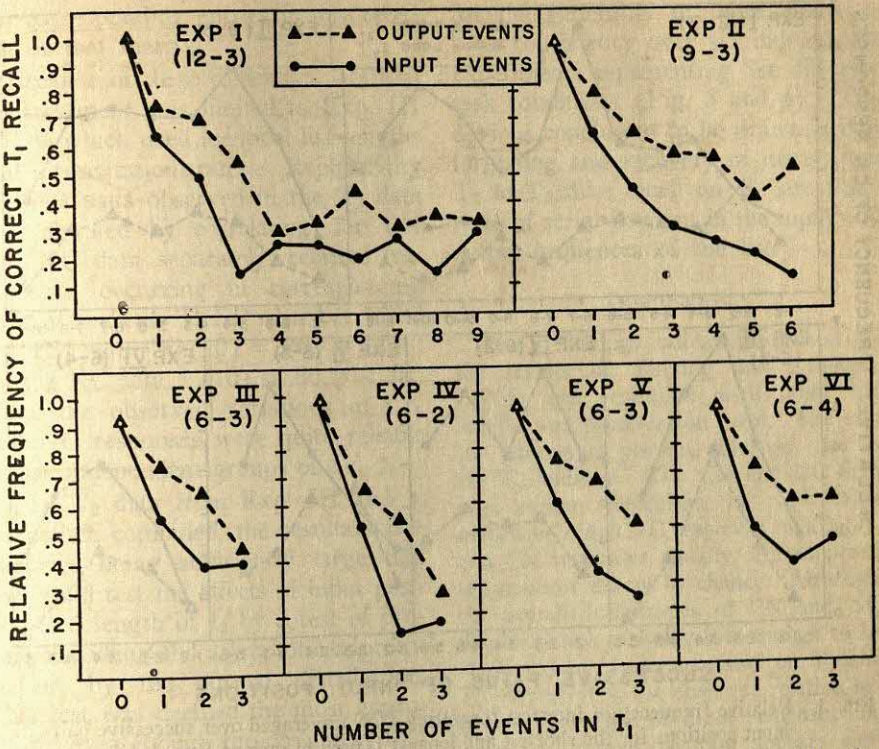


FIG. 2. Relative frequency of correct recall on T_1 for Exp. I–VI as a function of length of I_1 and kind of interpolated activity.

Figures 3 and 4 illustrate the relative frequencies of CN and NC responses, respectively, as a function of input position for the shortest and longest I_2 used in each experiment (short $I_2 =$ 1 interpolated output in all experiments; long $I_2 = 13$, 9, and 5 interpolated outputs in Exp. I, II, and III–VI, respectively). The primacy effect was again evident, early inputs being less frequently forgotten and more frequently recovered than middle inputs. Late inputs also were generally less likely to be forgotten and, with the long I_2, more likely to be recovered, than were middle inputs.

Statistical evaluation of the CN and NC data was complicated by two factors: (*a*) their absolute frequencies in each experiment were small, particularly when broken down over the $L^2/3$ conditions of the T_2 SP matrices, and (*b*) almost invariably, CN and NC entries in any one cell of a T_2 SP matrix did not involve the same *S*s or even the same number of *S*s as did

TABLE 1

ABSOLUTE (AF) AND RELATIVE (RF) FREQUENCIES OF OCCURRENCE OF FOUR CLASSES OF T_2 RESPONSES

Exp.	Response Class							
	C_1C_2		C_1N_2		N_1C_2		N_1N_2	
	af	rf	af	rf	af	rf	af	rf
I (12-3)	533	.22	170	.07	186	.08	1511	.63
II (9-3)	361	.27	108	.08	115	.08	766	.57
III (6-3)	216	.36	62	.10	45	.08	277	.46
IV (6-2)	141	.23	66	.11	46	.08	347	.58
V (6-3)	194	.32	71	.12	50	.08	285	.48
VI (6-4)	253	.42	43	.07	41	.07	263	.44

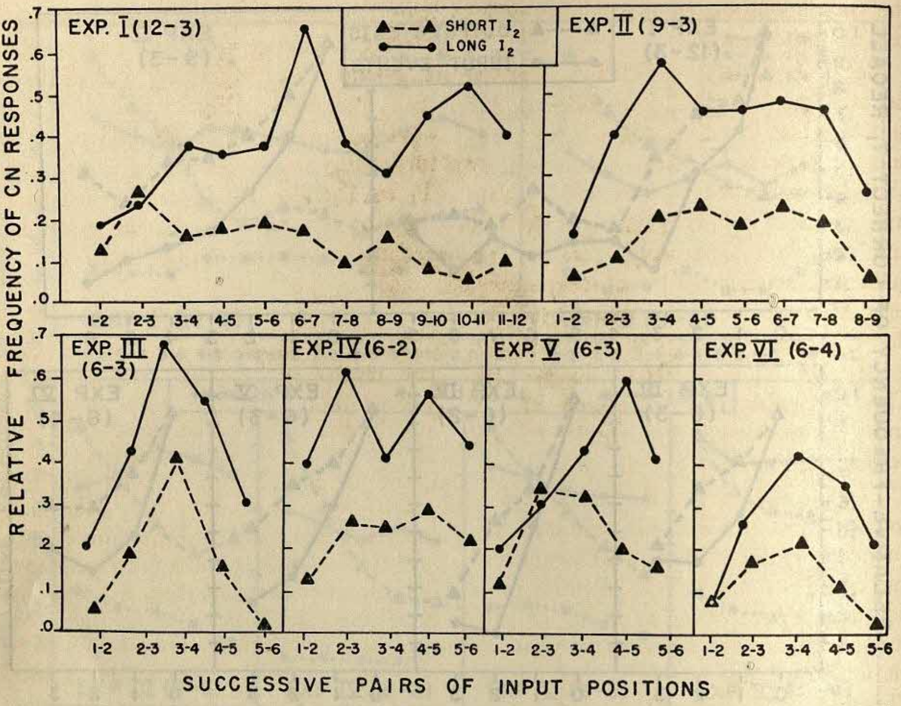


FIG. 3. Relative frequency of intertest forgetting (CN) averaged over successive pairs of input positions for the shortest and longest I_2 used in each of Exp. I–VI.

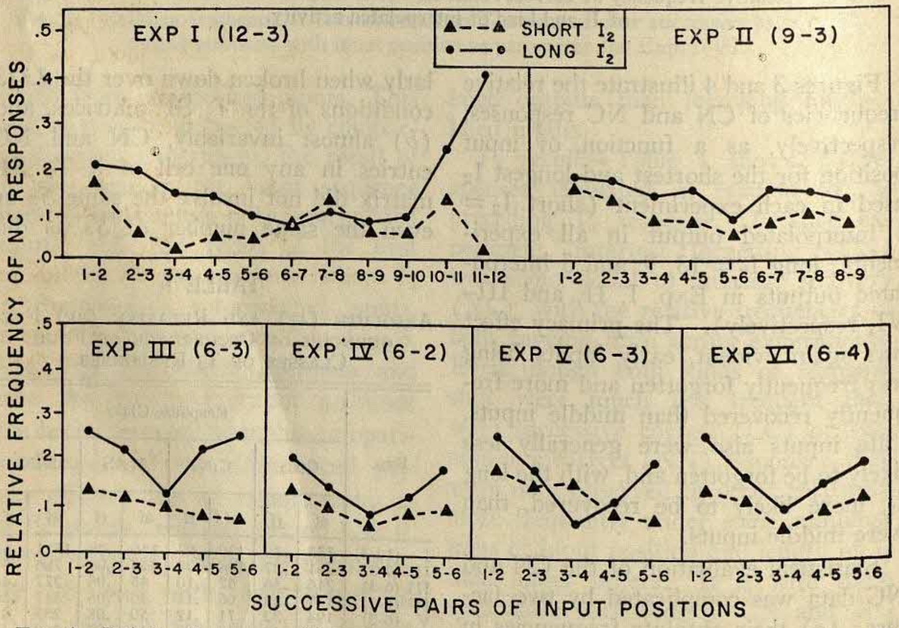


FIG. 4. Relative frequency of intertest recovery (NC) averaged over successive pairs of input positions for the shortest and longest I_2 used in each of Exp. I–VI.

the corresponding entries in any other cell of that matrix.

Because of these difficulties, statistical treatment was limited to Exp. III and V which used identical list lengths and presentation rates. Replicability of the trends observed in the T_2 data was checked by correlating, for CN and NC data separately, relative frequencies occurring in corresponding cells of the two T_2 SP matrices. For the CN data, $r\ (10) = .76$, $p < .01$; for the NC data, $r\ (10) = .86$, $p < .01$. Thus the observed variations in CN and NC responses were quite reliable across independent groups of *S*s.

The T_2 data from Exp. III and V were then combined, the resultant frequencies being sufficiently large that one could test the effects of input position and length of I_2 by a test of proportions using the asymptotic approximation by the normal distribution. This test was deemed the most appropriate, even though the possibility that the proportions were not completely independent was recognized.

Both forgetting and recovery, averaged across input positions, were significantly greater with the long I_2. For CN responses, $z = 5.19$, $p < .001$; for NC responses, $z = 2.37$, $p < .05$. The effects of input position, averaged across pairs of input positions, were tested separately for the short and long I_2. As Table 2 shows, early and late inputs were significantly less likely to be forgotten than were middle ones, regardless of length of I_2. For the short I_2, recovery was significantly more likely for early inputs than for late ones, while for the long I_2, both early and late inputs were more likely to be recovered than were middle ones.

While the statistical analysis thus indicates that input and output position did affect T_2 recall in Exp. III and V, probably the more impressive evidence for the generality of these effects is their consistency over six independent experiments representing five different task conditions (Fig. 3 and 4). The obvious conclusion to be drawn is that forgetting and recovery of items from T_1 to T_2, like recall on T_1, are functions of serial position in the input and output sequences of the list.

DISCUSSION

The T_1 findings indicate that the original results of Tulving and Arbuckle (1963) are replicable with other list lengths and presentation rates. The major interest at present, however, lies in the T_2 findings. The CN and NC data pose serious difficulties for all-or-none models in which NC responses invariably, and CN responses usually, are supposed not to occur except by chance. Although the overall frequencies of CN and NC responses are probably low enough to be attributed to "lucky guesses" on T_1 and T_2, respectively, the observed relation between occurrence of these responses and input and output position appears contrary to any guessing interpretation.

Are the observed patterns of T_1 and T_2 recall produced by differences in the learning of items, differences in their re-

TABLE 2

SUMMARY OF z VALUES OBTAINED IN TESTS OF PROPORTIONS OF CN AND NC RESPONSES OCCURRING UNDER DIFFERENT COMBINATIONS OF INPUT POSITION AND LENGTH OF I_2

Length of I_2	Input Position Comparisons		
	Early[a] vs. Middle	Late vs. Middle	Early vs. Late
CN			
Short	4.22***	3.17**	.77
Long	4.43***	2.89**	1.88
NC			
Short	.94	1.66	2.39*
Long	2.63**	2.34*	.23

Note.—Data from Exp. III and V combined.

[a] "Early" represents data from Input Positions 1 and 2, "middle," from 3 and 4, and "late," from 5 and 6.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

tention, or both? It seems paradoxical that, while the one-trial learning controversy revolved around whether nonrecall of items meant "no learning" or "insufficient learning," the present data, like those of Melton (1963), Murdock (1961), Peterson (1963), and Waugh and Norman (1965) cited above, indicate that neither of these "negative states" exists. If a single pair of well-integrated verbal items is tested immediately after presentation, the B member can always be recalled, though it may not be recalled if testing is delayed (e.g., Fig. 2). The most reasonable explanation of these data is that learning is always *all* and that whether or not a given item is recalled depends solely on whether or not it has been forgotten during I_1 and I_2.

The primacy effect, though, would seem to make any such refutation of differential learning more difficult. The incremental theorist could argue, following Underwood (1964), that the primacy effect reflects differences in response strength which, immediately after presentation, are supraliminal and hence not measurable. The counterargument would be that all items attain 100% response strength on one presentation, but they are forgotten at different rates depending on factors such as input position. Bernbach (1965) has expressed such a conception of learning as differential forgetting in a stochastic model, and has shown that this model fits PA data from a variety of experiments quite well.

If one accepts this view of learning as differential forgetting, then one must account for why such forgetting occurs. The finding that recall on T_1 and T_2 depends on input position and T_1 and T_2 output position indicates that factors operating within the learning trial produce this forgetting. The finding that a retention interval filled with outputs produces less forgetting than does the same retention interval filled with inputs strongly suggests that it is the activity engaged in by *S* and not the mere passage of time which causes items to be forgotten. This particular finding has also been reported by Tulving and Arbuckle (1966) under conditions where number of interpolated inputs and input position were not confounded as they were here.

This explanation of why items are not recalled on a learning trial is essentially an intralist-interference explanation. It can be extended to the recovery data also, with the greater number of NC responses for the long I_2 being attributed to the fact that the longer time interval permitted the dissipation of specific interference effects present at T_1.

While intralist interference was clearly present, there was no evidence for interlist interference. If anything, performance of *S*s improved over successive lists. Furthermore, even though all experiments used what was essentially an A-B A-C design across lists, only 3 of the 18,450 T_1 responses were B members from previous lists. Apparently, the selection of all B members for a given list from the same conceptual category was sufficient to eliminate interference from previous lists.

REFERENCES

ARBUCKLE, T. Y. *Intratrial interference in immediate memory.* (Doctoral dissertation, University of Toronto) Ann Arbor, Mich.: University microfilms, 1964, No. 66-1061.

BATTIG, W. F., & BRACKETT, H. R. Comparison of anticipation and recall methods in paired-associate learning. *Psychol. Rep.*, 1961, **9**, 59-65.

BERNBACH, H. A. A forgetting model for paired-associate learning. *J. math. Psychol.*, 1965, **2**, 128-144.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207-233.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1-21.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618-625.

Peterson, L. R. Associative memory over brief intervals of time. *J. verbal Learn. verbal Behav.,* 1963, 2, 102–106.

Tulving, E. Intra-trial and inter-trial retention: Notes toward a theory of free recall verbal learning. *Psychol. Rev.,* 1964, 71, 219–237.

Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. *J. verbal Learn. verbal Behav.,* 1963, 1, 321–334.

Tulving, E., & Arbuckle, T. Y. Input and output interference in short-term associative memory. *J. exp. Psychol.,* 1966, 72, 145–150.

Underwood, B. J. Degree of learning and the measurement of forgetting. *J. verbal Learn. verbal Behav.,* 1964, 2, 112–129.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychol. Rev.,* 1965, 72, 89–104.

(Received July 25, 1966)

(Continued from page 442)

Speed-Accuracy Tradeoff in Reaction Time: The Effect of Discrete Criterion Times: Robert G. Pachella and Richard W. Pew*: Department of Psychology, Human Performance Center, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

Relative Desirability or Aversiveness of Immediate and Delayed Food and Shock: K. Edward Renner* and Laura Specht: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

An Examination of Recognition and Free Recall as Measures of Acquisition and Long-Term Retention: Darryl Bruce and Charles N. Cofer*: Department of Psychology, 117 Burrowes Building, Pennsylvania State University, University Park, Pennsylvania 16802.

An Experimental Analysis of Single Stimulus Tests and Multiple Choice Tests of Recognition Memory: Walter Kintsch*: Department of Psychology, University of California, Riverside, California 92502.

Inference and Prediction of Events and Event Generators: Alan J. Miller*: Department of Psychology, University of California, Irvine, California 92650.

*Asterisks indicates author for whom address is given.

Journal of Experimental Psychology
1967, Vol. 74, No. 3, 452–454

NONSTIMULUS AND NONRETINAL MECHANISMS IN FIGURAL AFTEREFFECTS[1]

EDWIN P. WILLEMS

Rice University

Ganz's (1966) recent attempt to explain figural aftereffects (FAEs) as a special case of simultaneous illusion is questioned on the grounds of incompleteness, i.e., Ganz's postulation of afterimage as the critical factor relating FAE to illusions does not permit derivation of some empirically demonstrated relationships. Data are reported suggesting that variation in set to perceive an ambiguous inspection figure produces FAE in as many Ss as does variation in the inspection figure itself.

Ganz (1966), attempting to synthesize the findings from studies of geometrical illusions and figural aftereffect (FAE), suggests that FAEs are simply a special case of simultaneous illusion. The present paper argues that Ganz's theory is incomplete; i.e., that there are FAE phenomena that cannot be derived from, or subsumed under, his principles.

Ganz concludes, ". . . it appears possible that most of the quantitative properties of the figural aftereffects will be reduced to the operation of a number of basic visual mechanisms: adaptation, inhibition, and ocular tremor [p. 148]," and his conceptual sequence may be schematized as follows: (*a*) FAE is a function of (*b*) the inhibiting effect of an inducing figure, i.e., afterimage of the inspection figure, upon the test figure. (*c*) This inhibiting effect occurs in the neural correlates of the inducing figure and test figure, which are a function of (*d*) the physically and objectively defined inspection and test figures. Ganz thus limits his principles to properties of the stimulus array, known and assumed variables of physical optics, and a great deal of emphasis upon afterimages and lateral inhibition.

Story (1959) reasoned that the cortical representation of a closed B is denser than the cortical representation of a broken B-13, and that the closed B as an inspection figure should result in more frequent FAE than the broken B-13. Story reasoned further that if *S*s received a set to perceive the broken B as a B, they would tend toward closure of the broken B and more frequent FAE than *S*s who received a set to perceive the broken B as the two-digit number, 13. Story's data supported her line of reasoning, and one interpretation is that FAE was a function of what *S*s *assumed* about the inspection figure. The mechanism of such an assumption, a function of set, is not derivable from Ganz's theory.

Method

Story's technique was extended in the present experiment. One modification involved arranging type of set (letter or number) and type of inspection figure (closed or broken B) into an orthogonal design. Half of the *S*s (letter set) viewed a series of cards on each of which there was a capital letter (K, F, R, A) drawn with gaps. The letters alternated from right to left of the fixation mark, and the fifth card had the inspection figure to the right of the mark. Half of the *S*s (number set) viewed a series of two-digit numbers (37, 44, 19,

---

[1] The author wishes to thank Donald H. Lucas, who helped with the gathering of data, and Theron Stimmel, Peter Leppman, and T. W. Wann, who read the manuscript and made helpful suggestions.

12) alternating from right to left of the fixation mark. The fifth card in the series was again the inspection figure. Half of each of these groups viewed a closed B as an inspection figure, while the other half viewed the broken B, yielding four main groups. There were two control groups, one of which received the letter set and the other the number set, but which, instead of viewing an inspection figure, viewed a card with only a fixation mark and no inspection figure.

The second modification of Story's procedure was to make set implicit, or intrinsic to the experimental task, rather than dependent upon verbal labeling. At the presentation of each card, including the card with the inspection figure, *S* was asked to report what the symbol was, without any suggestion by *E*.

Figures were drawn in India ink on white cards 9½ × 22 in. in overall size. All letters and numbers were drawn 2-in. square, with lines ¼-in. thick. Figures were 3½ in. from fixation marks. All *S*s viewed the same test figure, composed of two vertical bars, each ¼-in. thick and 1½-in. high, placed equidistant from the fixation mark so that the right-hand bar would have fallen in the gap of the broken B of the inspection figure. The *S*s viewed all figures binocularly, with their heads on a chin rest located 3½ ft. from the card holder. Homogeneous light brown paper covered the wall 1 ft. behind the card holder and the table top.

Fifty-six students from an introductory psychology course participated, 10 for each of the four main groups and 8 for each of the two control groups. After practice for focusing on a fixation mark, *S*s were told that they would be looking at a series of cards with symbols on them and were then presented one of the two sequences leading up to the inspection figure. After reporting what he saw on the inspection card, *S* fixated the inspection card for 120 sec. Reminded not to move his head, *S* was then immediately shown the test figure, and was asked progressively: "Does anything strike you? Does anything strike you about those two bars? Do the bars differ in any way? Do they differ in size? Is one closer than the other? Is one brighter?" A response was scored as positive FAE if *S* reported that the right-hand bar of the test figure was smaller, dimmer, or farther away than the left-hand bar, or that the left-hand bar was larger, brighter, more vivid, or closer. Other outcomes, including the reverse of positive FAE and no difference between the bars, were recorded.

TABLE 1

Number of *S*s Reporting Positive FAE, No Difference between the Bars, and Reverse FAE

Group	*N*	Set	Inspection Figure	+FAE	No Difference	Reverse FAE
1	10	letter	closed	6	4	0
2	10	letter	broken	9	1	0
3	10	number	closed	9	1	0
4	10	number	broken	4	4	2

Results and Discussion

A chi-square comparing the frequencies of *S*s reporting FAEs in the two control groups (letter vs. number set, but no inspection figure) showed no difference in effects of the two set sequences alone. There was a three-way split among the 16 *S*s in the control groups, with 5 reporting positive FAE, 5 reporting no difference between the bars, and 6 reporting the reverse of positive FAE. The results for the four main experimental groups are shown in Table 1. A 2 × 2 multivariate chi-square (Sutcliffe, 1957) indicated that (*a*) the effect of the closed inspection figure (Groups 1 and 3) did not differ from the broken inspection figure (Groups 2 and 4) across types of set sequence, $\chi^2 = .476$, and (*b*) effect of letter set (Groups 1 and 2) did not differ from number set (Groups 3 and 4) across types of inspection figure, $\chi^2 = .476$. (*c*) However, inspection figure and set sequence interacted significantly, $\chi^2 = 7.619$, $p < .005$. This analysis was based upon a division of *S*s into those who reported positive FAE and those who did not. Inspection of Table 1 suggests that in comparison with the other groups, Group 4 (number set, broken B inspection figure) produced the fewest positive FAEs. In fact, the only reversals from positive FAE in the experimental groups occurred in this group. Perhaps most important, letter set with the broken B inspection figure (Group 2) produced as high a frequency of FAE as the closed B inspection figure (Groups 1 and 3). One interpretation of these findings, which corroborate Story's, is that when the inspection figure could be perceived in two different ways, what *S*s assumed about the inspection figure determined FAE.

It should be made clear at what point Ganz's theory of FAE is being questioned. In making FAE a special case of simultaneous illusion, Ganz suggests those mechanisms and processes of which FAE is a function. The crucial step in his argument involves the mechanism of afterimage of the inspection figure, with its processes of light and dark adaptation and inhibition of test figure contours. He says:

> The afterimage exists beyond the preservation of the distal figure and decays in time. This is the mechanism which bridges the time gap in the figural aftereffect situation; or more properly, the processes of light and dark adaptation bridge that time gap [p. 133].

Ganz thus purports to have explained the phenomena of FAE, including the proaction of the inspection figure. However, afterimages are usually assumed to be a function of two sets of factors: (*a*) excitation and adaptation characteristics of the visual receptors and (*b*) the stimulus array. Thus, as Ganz suggests by his review of relevant literature, one way to assess the tenability of the crucial step in his theory is to vary the known parameters of stimulus array and afterimage and observe their effects upon FAEs. However, in the study by Story and in the present experiment, these were precisely the variables held constant for certain *S*s, while what they were led to *assume* about the inspection figure was varied. How *S*s perceived the inspection figure was found to influence FAEs for as many *S*s as the variation in the inspection figure, i.e., the stimulus array, itself. Therefore, if the question is, "Of what is FAE a function?" then Ganz's theory is incomplete. Stated another way, Ganz's theory, as well as the theories of Köhler and Wallach (1944) and Osgood and Heyer (1952), attempt to account for FAE in terms of stimulation pattern and properties of the visual receptors and visual cortex, but none of the theories takes into account the kind of nonstimulus, nonretinal, i.e., cognitive, factors that appear to be operating in the data reported here.

REFERENCES

Ganz, L. Mechanism of the figural aftereffects. *Psychol. Rev.*, 1966, **73**, 128–150.

Köhler, W., & Wallach, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.

Osgood, C. E., & Heyer, A. W. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

Story, A. Figural after-effects as a function of the perceived characteristics of the inspection-figure. *Amer. J. Psychol.*, 1959, **72**, 46–56.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

(Received July 27, 1966)

Journal of
Experimental Psychology

Vol. 74, No. 4
Part 1 of 2 Parts

August 1967

EFFECTS OF INDUCED CHUNKING ON TEMPORAL ASPECTS OF SERIAL RECITATION [1]

R. S. McLEAN AND L. W. GREGG

Carnegie Institute of Technology

75 Ss learned serial lists composed of random orders of 24 letters of the English alphabet. The main independent variable was stimulus grouping, where the letters were presented either 1, 3, 4, 6, or 8 letters at a time. After reaching the criterion of 1 perfect recitation, Ss were required to recite the serial list backward. Time intervals between the letters during the criterion and backward recitations were analyzed to show the induced chunking produced by the stimulus grouping.

"Chunking" has remained largely a descriptive concept since it was first introduced by Miller (1956). Although it is used extensively in explanations of verbal learning this concept of higher organization of the stimuli in memory has rarely been objectified to the point of permitting the identification and measurement of structure in verbal performance. Tulving (1962) proposed a measure of organization based on information-theoretical concepts applicable to the free-recall paradigm where all pairs of *n* stimulus items are presented over the course of *n* trials. The measure is based on the occurrence of ordered pairs in free recall and yields a subjective organization score which characterizes a particular task or group of *S*s, much as the traditional learning curve does. To characterize the processes used by a single *S* in a verbal learning situation, one must be able to examine the organization imposed by *S* on the task materials.

The present study makes explicit one way in which structural characteristics of *S*'s encoding can be derived in the case of serial verbal learning. The approach is through an analysis of the temporal properties of *S*'s ordered recall of the serial list. In examining these properties, groups of items recited together quickly are assumed to form a coherent memory unit or "chunk." Longer or shorter intervals between chunks may indicate different kinds of encoding relationships, rather

---

[1] This research was supported by the Public Health Service by Grant MH-07722. The work was completed while R. S. McLean held a predoctoral research fellowship, 1-F1-MH-29, 112-01, also from the Public Health Service.

than simply differences in the strengths of bonds between pairs of elements.

The organization developed by *S* must result from the interaction of the actual stimuli, the environmental events, and the information processing strategies of *S*. Each of these may contribute to the development of chunks. The following serve as examples. (*a*) Some stimuli may already form a unit with which *S* is familiar. (*b*) External punctuation of the stimuli may serve to create groupings of the individual elements. (*c*) The *S* may monitor his own performance and impose structure by selective attention, rehearsal, or other means.

In the present experiment, the second of these techniques, specifically spatial grouping of letters of the English alphabet, was used to impose external structure upon the serial list. It was hypothesized that the temporal structure in ordered recall—and hence the inferred structure of *S*'s internal representation—would reflect both the number and size of the letter groups. Immediate backward recitation of the just learned list was required of each *S*. This procedure was adopted from a report by Allan (1961) on the supposition that the effect of induced chunking would be even more pronounced.

METHOD

Stimuli and experimental design.—Letters of the English alphabet, excluding E and I, were randomly permuted to form five different 24-letter lists. The lists were:

List 1—UARSKYCNLHOXTPJGDFQZBWVM
List 2—MGVSDLJYRFAKBTCWQOPXUNHZ
List 3—DWUCJZXOGSPBRVYLMKATHNQF
List 4—CRNTYMPVFHSKXAWJZUQLDOGB
List 5—UPRKGQXYZSFDHVBCMJLANOWT

The letters were typed on 3 × 5 in. file cards to form groups of 1, 3, 4, 6, or 8 letters on a card. The design of the experiment was a 5 × 5 factorial with three *S*s per cell.

Subjects.—Seventy-five undergraduate college students served individually in the experiment to fulfill a course requirement. Each *S* received one list in one grouping condition, assigned randomly.

Procedure.—The experiment was self-paced. In the instructions, *S* was asked to turn over the cards, one at a time, reading aloud the letters that appeared on that card. The *S* was asked to enunciate each letter distinctly and to maintain a uniform rate while reading. After each pass through the deck, *S* attempted to recall the list in the correct order. The read-recall trials were continued to a criterion of one perfect recitation. When this criterion was reached, *S* was requested to say the entire list in reverse order; *S* was not instructed about the backwards trial until criterion had been reached.

The experimental sessions were recorded on magnetic tape; these recordings were used to obtain the time of occurrence of each verbal response on the criterion trial and the backward recall trial. Timing was accomplished by use of an *O* who listened to the tapes and indicated the occurrence of each letter by pressing a push-button switch. Presence of verbal activity at the time a response was indicated by *O* was verified by a voice key. Data recording was accomplished automatically through the facilities of the Computer Controlled Psychological Laboratory. Two *O*s were used in scoring the records. Inter- and intra-*O* measures indicated that the time values were accurate to about 5% for records of those *S*s who responded most rapidly. The percentage of error fell to less than two for records of longer duration, 15 sec. or greater.

RESULTS

Table 1 presents the group mean total learning times as a function of stimulus grouping. An analysis of variance showed a significant effect of stimulus grouping on the total time required to learn the list of letters, $F\ (4, 50) = 8.23$, $p < .001$. Time to learn also varied over lists, $F\ (4, 50) = 4.05$, $p < .01$, but the Grouping × Lists interaction was not significant, $F\ (16, 50) = 1.45$.

The extent to which grouping of responses in recall paralleled the grouping of stimuli was tested by examining the interletter response times of *S*s as a function of "perfect" chunking pat-

TABLE 1

Group Mean Total Learning Time in Seconds

List	Stimulus Grouping					
	1	3	4	6	8	*M*
1	772	486	692	387	252	518
2	1054	366	310	602	425	551
3	865	555	539	441	618	604
4	606	355	202	292	461	383
5	464	310	364	397	293	366
M	752	414	421	424	410	484
SD	303	156	232	205	180	

terns. The 23 interletter times were designated "interchunk" or "intrachunk" times depending upon their serial location within the list according to the chunking pattern being tested. For example, if a "three's" pattern were being tested, the first and second interletter times would be considered to be intrachunk times and the third interletter time would be considered an interchunk time. For each pattern to be tested each *S*'s interletter time was summed in its appropriate category. A chunking pattern score was then computed by forming a weighted ratio of interchunk to intrachunk times. If the effect of the stimulus grouping is present in the responses, the score for the pattern that corresponds to the stimulus grouping should be greater than unity and should also be greater than the other scores.

Table 2 summarizes the mean chunking pattern scores for the five patterns tested. In each column the largest chunking pattern is obtained when the tested pattern is equivalent to the stimulus grouping. An analysis of variance for each individual pattern showed significant differences in the scores attributable to the stimulus grouping. In one case, the interaction of lists and groupings was significant ($p < .05$).

A measure of grouping independent of any particular external grouping may be obtained by dividing the letters into groups separated according to decreasing interletter times such that interletter times within the group are smaller than those between the groups. The smallest groupings were obtained which do not isolate a single letter as a "group." These groupings provide information about the regularity of the verbalizations of *S*s. A test of the uniformity with which groupings of the same size were detected within the criterion recall responses was made by computing the standard deviation of grouping size for each *S*. Table 3 gives these results. An analysis of variance showed that stimulus grouping affected the degree of variability in emitted groups, $F\ (4, 50) = 2.74,\ p < .05$; the lists did not contribute to that variability, $F\ (4, 50) = .46$.

The effect of stimulus grouping on memory structure is also seen in its effect on grouping in the backwards recall. One way to detect the effect is to apply the chunking-pattern measure employed on the forward criterion trial data. Table 4 summarizes the mean chunking-pattern scores for five hypothetical patterns on the backward recitation data. These scores are somewhat diluted since there was no guarantee that exactly 24 letters would be recalled by *S*. If *S* recalled additional

TABLE 2

Summary of Chunking Patterns for the Criterion Trial

Pattern Tested	Stimulus Grouping					
	1	3	4	6	8	*M*
2	1.28	1.20	4.35	2.18	2.58	2.32
3	1.61	8.14	.98	4.69	.68	3.21
4	1.25	.79	6.29	1.86	3.26	2.69
6	1.53	3.79	1.89	5.71	.98	2.78
8	.92	.33	3.58	1.77	4.70	2.26

Note.—Strategy tests are based (in the case of the criterion trial) on data that has been adjusted to a uniform 24 letters per *S*; additional letters were removed. Individual analyses of variance on each pattern test showed significant effects of the stimulus grouping ($p < .01$) on all tests, a significant effect of the interaction of stimulus grouping with stimulus list in one of the five tests (3's test, $p < .01$).

TABLE 3

UNIFORMITY OF GROUPING FOR THE CRITERION TRIAL

List	Stimulus Grouping					
	1	3	4	6	8	*M*
1	4.01	2.18	.96	.81	2.72	2.13
2	1.57	0.00	.33	4.91	1.29	1.62
3	3.16	1.52	.60	1.53	1.46	1.65
4	3.86	.47	.70	2.23	2.32	1.92
5	3.81	1.41	3.39	2.13	1.77	2.50
M	3.28	1.12	1.19	2.32	1.91	1.97

Note.—Mean standard deviations of size of chunks identified by smallest multiletter grouping criterion. Each cell is mean for three cases.

letters or omitted letters, the scoring algorithm would not necessarily split the response list into ideal groups corresponding to those groups found in the stimulus list under the same pattern, thereby reducing the strategy score. In spite of this limitation, the scores show strong effects of the stimulus grouping. Individual analyses of variance for each test showed significant effects of stimulus grouping ($p < .01$); there were no other significant effects.

The patterns in forward and backward recall may be examined for similarity by correlating the interletter times between corresponding letters on the two trials. A separate coefficient was computed for each *S*. These correlations were converted to z scores and an analysis of variance showed significant effects of both stimulus grouping, $F\ (4, 50) = 3.09$, $p < .05$, and lists, $F\ (4, 50) = 3.33$, $p < .05$. Forty-seven of the 75 individual correlation coefficients were significant beyond the .05 level; all but 3 of the remaining 28 coefficients were positive.

TABLE 4

SUMMARY OF CHUNKING PATTERNS FOR THE BACKWARDS TRIAL

Pattern Tested	1	3	4	6	8	*M*
2	1.23	.78	5.29	1.69	2.65	2.33
3	1.44	9.55	.81	4.41	.87	3.42
4	1.28	.91	7.44	1.12	3.91	2.93
6	1.26	2.38	1.43	4.09	.85	2.00
8	1.37	.51	3.97	.43	5.48	2.35

Note.—Pattern tests are based (in the case of backwards trials) on complete performance data; the number of letters spoken will vary between *S*s. Individual analyses of variance on each strategy test showed significant effects of the stimulus grouping ($p < .01$) and nonsignificant effects due to individual lists of stimuli and the interactions of lists with stimulus grouping.

TABLE 5

MEAN RECITATION TIMES IN SEC. FOR FORWARD (CRITERION) AND BACKWARD TRIALS

	Stimulus Grouping					
	1	3	4	6	8	*M*
Forward (F)	32.4	27.8	18.1	23.2	19.8	24.2
Backward (B)	69.2	47.8	45.9	48.8	43.5	51.0
Ratio (B/F)	2.2	1.7	2.5	2.2	2.2	2.1

With respect to the absolute rates of responding, *S*s in this self-paced task required about twice as much time for the backward recitation as for the final trial during acquisition. In Table 5, mean recitation times and the ratios of the times for backward and forward recitation are shown.

Only 26 of the 75 *S*s were able to recite the serial list backwards without errors. The mean errors for the grouping conditions were 1.07, 3.26, 2.80, 1.60, and 2.80 for the stimulus groupings of Sizes 1, 3, 4, 6, and 8, respectively. An analysis of variance of the raw errors indicated that neither of the main effects of stimulus grouping or lists nor the interaction was significant.

DISCUSSION

The results of the several analyses provide evidence that stimulus grouping leads to stable chunks of the size induced by the grouping. The chunks are stable because there is good correspondence between the temporal aspects of the criterion trial and the backward recall (Table 4). Whatever the underlying structure of *S*s internal representation, that same structure was exhibited in the modified task, The mean chunking-pattern scores are maxima in those cells of Table 2 for corresponding stimulus grouping (except for the one-at-a-time condition). This fact coupled with the uniformity observed in the size of the groupings (Table 3), particularly for stimulus groupings of Sizes 3 and 4, confirms the effectiveness of the experimental variable in inducing chunks of a particular size. These effects are generalizable over the lists used in this experiment.

The *S*s who received the items one at a time also produced stable chunks, but chunk size varied. There was a tendency for these *S*s to group by three's; 24 of the 75 chunks produced by the 15 *S*s in this experimental condition were of Size 3. This was the modal frequency; the mean of the distribution was 4.8. However, no two *S*s produced precisely the same chunking patterns in the one-at-a-time condition. Eight of the 15 *S*s in the three's condition and four of the 15 *S*s in the four's condition did have identical patterns.

The preponderance of the evidence suggests that no single feature intrinsic to any of the serial lists produced unique encodings of the 24 letter strings. The chunking patterns of *S*s are easily manipulated by the stimulus grouping imposed by *E*. That operation, independent of the particular letter sequences, far overshadows any other perceptual or associational relationships; and there were many such relationships within these randomly selected lists. No effort was made to eliminate them. There were, for example, embedded words—"sky" in List 1, "at" in List 3, "Dog" in List 4, and two words, "up" and "now" in List 5. List 5 was learned significantly faster on the whole than the other lists.

It would appear, then, that information both from features of the lists themselves and from the manipulation of stimulus grouping induced chunking in *S*s organization of a serial list. It seems likely that this organization takes on a hierarchical form. At the top level of the hierarchy are those cueing features that allow *S* to get from one chunk to another. At a lower level, within chunks, additional cues enable *S* to produce the integrated strings that become his overt verbal responses.

While the distinction between paired-associates learning and serial learning is a useful one for describing *E*'s experimental procedures, *S*'s behavior in any learning task rests upon internalized sets of relationships that are neither purely paired associates nor serial motor in form. Particularly for those *S*s in grouping conditions of Sizes 6 or 8, where there were only three or four separate 3 × 5 in. cards, the obvious pairing of serial position name and card label occurred. During acquisition and during the backward recitation many of these *S*s referred to the "first card," "last card," etc., and were able, then, to state a specific fact about the content of the card, e.g., the first or last letter, or an idiosyncratic feature of the letter sequence on that card. Having once "located" a desired card in memory, *S* might then produce a completely integrated response, or he may have to search further for the subunits representing the serial-motor level of the hierarchy (e.g., "first half," "last half").

Thus conceived, the laws of learning cannot be different for serial and PA learning tasks. In fact, the distinction between response integration and association learning in the PA paradigm (Underwood & Schulz, 1960) is of little importance. What does become important is the discovery of those properties of the human information processing system that allow encoding and retrieval of the source materials in such flexible and efficient ways.

REFERENCES

ALLAN, M. D. Memorizing, recoding, and perceptual organization. *Brit. J. Psychol.*, 1961, **52**, 25–30.

MILLER, G. A. The magical number seven, plus or minus two: Some limits in our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–354.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1963.

(Received August 29, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 460–465

STRESS EFFECTS ON SKILL[1]

M. PAUL WILLIS

Montana State University[2]

During 60 1-hr. acquisition trials on a rifle-marksmanship task, 20 Ss (high-school males) encountered low stress (2 ft. lb. of recoil) while another 20 Ss encountered high stress (25 ft. lb.). Subsequently, all Ss received 24 transfer trials, ½ of each group operating under unchanged stress levels and ½ under reversed stress levels. The results suggest that stress affects transfer performance in 3 distinguishable ways: (a) through stimulus generalization, (b) through acquisition skill level, and (c) directly, i.e., independently of stimulus generalization and acquisition skill level.

Experimental investigations of the effects of stress upon human performance have frequently been criticized in terms of deficiencies in the particular stress variable selected. Lazarus, Deese, and Osler (1952) pointed out the paucity, if not complete absence, of studies utilizing stress variables which adequately meet the following criteria: (*a*) realism, (*b*) amenability to experimental manipulation, and (*c*) independence from *S*'s ability to learn. The present investigation represents an attempt to study the acquisition and transfer of motor skill in a task which provides a stress variable meeting these criteria.

Deese (1962) suggests that approaching the problem of stress in training by means of the transfer paradigm is a reasonable approach, particularly if one accepts the identification of the concept of stress with stimulus conditions. He suggests further that, despite the general usefulness of the concept of stimulus generalization, other sources of differences between training and performance (transfer) situations "... will be so much more important as to swamp any effects due to generalization decrement [p. 219]."

The psychomotor task adopted for the present study involved big-bore, or centerfire, rifle target shooting. Stress was defined in terms of the physical stimulus variable, ft. lb. of recoil. This stress variable is: (*a*) realistic—it is not artificially imposed by *E* but rather is intrinsic to, or inherent in, the task; (*b*) amenable to experimental manipulation over a rather wide continuum—from near zero to actually hazardous levels; (*c*) one which permits a measure of stress effect on performance which is independent of skill level or learning. The third criterion was met by incorporating a "bench rest" measure of performance, considered by such authorities as Whelen (1954, 1957) to be highly sensitive to stress level but relatively insensitive to learning or skill level.

METHOD

Subjects.—The *S*s were 40 experimentally naive, right-handed, male high-school students of normal vision and body build.

Equipment.—Remington "Model 700" bolt-action rifles chambered for .222 Remington centerfire and .308 Norma Magnum calibers, equipped with 15-power telescopic sights

---

[1] This research was supported by United States Public Health Service Grant MH-04215-02 from the National Institute of Mental Health. Appreciation is due also to Earl L. Seiler and Jerome V. Lisovich for numerous contributions during the course of this research.

[2] Formerly at American Institutes for Research.

(Unertl 2 in. Ultra Varmint), and adjustable target slings. Accessory equipment included shooting jackets, gloves, safety glasses, ear plugs, shooting mats, bench rests, and cartridges handloaded to produce specific recoil levels. Targets presented had 1 in. bench-rest squares (aiming points) and were located 50 meters from the firing point.

Procedure.—The 40 *S*s were randomly divided into two groups of 20 *S*s each for acquisition training, one group (Low stress) encountering approximately 2 ft. lb. of recoil while the other group (High stress) encountered approximately 25 ft. lb. of recoil. Each of these two groups was further divided (randomly) into subgroups of 10 *S*s each for transfer training, one subgroup in each case being switched to the previously unexperienced stress level while the other subgroup continued under unchanged stress conditions. For convenience in discussion, the four subgroups will be referred to as Low-Low (i.e., low stress during acquisition, followed by low stress during transfer), Low-High, High-High, and High-Low.

All *S*s were tested under both classical prone and modern bench-rest shooting positions for each of 15 blocks of acquisition trials and 6 subsequent blocks of transfer trials. At each trial block, bench (B) and prone (P) positions were presented in counterbalanced order (BPPB). For each trial block, two bench and two prone scores were obtained. Each score, whether bench or prone, represented the horizontal distance in sixty-fourths of an inch between extreme left-most and right-most shots in a series of five consecutive, slow-fire (spaced) shots. All scores were corrected by subtraction of a base-line value representing the maximum potential of the particular rifle and load as periodically estimated from median values of accuracy tests made by an expert competitive marksman firing under optimum bench-rest conditions. This procedure controlled for such factors as barrel wear and differences in individual rifles and loads. Each trial block consisted of a 1-hr. session. Only one session per day was scheduled for a given *S*.

The bench position, as previously indicated, was incorporated in the procedure in order to provide a measure of stress effect which was relatively independent of learning or skill level. In the bench position, the rifle was not held by *S* but was supported on special sand bags, cradle, and bench-rest table. Hence, in this position, holding or steadiness errors were ruled out. Aiming errors were also negligible in view of the 15-power target scope sights, the short range (50 meters), and the normal vision requirement. The bench situation, however, was considered to be sensitive to "flinching," i.e., anticipatory avoidance reactions, trigger "jerking," and other indications of "recoilitis," or sensitivity to recoil.

The conditions obtaining in the prone position were the same as for the bench condition with the major exception that the prone condition required *S* to support the rifle while lying prone on a shooting mat, with only a target sling and glove as steadiness aids. In other words, the prone position required *S* to hold or steady the rifle by means of proper positioning (orientation to target, sling adjustment, elbow placement, etc.). Each *S* fired only one string of five shots on a given target. The target scope sights were adjusted to produce bullet impact on the target below the aiming point, thereby leaving the aiming point undamaged during the string of five shots. Bullet impact point was clearly visible through the scope sight. Targets were scored immediately after a given 1-hr. session and scores were posted for *S*s' inspection on a daily basis. During the acquisition phase, *S* did not know which transfer stress level (i.e., same or different) he would subsequently encounter. Each *S* fired a total of 300 individual shots during the acquisition phase and 120 shots during the transfer phase.

RESULTS AND DISCUSSION

Mean error scores are plotted, in Fig. 1, as a function of acquisition and transfer trial blocks with position and stress conditions as parameters. Random assignment of *S*s to the subgroups resulted in two High-stress groups, and two Low-stress groups, with similar scores at the end of the acquisition trials. On the last acquisition trial, the prone means for the High-High and High-Low groups are 95.3 and 110.2, respectively, $t(18) = 0.51$, $p > .60$. Similarly, on the last acquisition trial, the prone means for the Low-Low and Low-High groups are 94.5 and 126.0, respectively, $t(18) = 0.96$, $p > .35$. Equality of the appropriate subgroups at the end of the acquisition phase is also reflected in the bench data. On the last acquisition trial, the bench means for High-High and High-Low groups are 62.0 and 81.2, respectively, $t(18) = 1.41$, $p > .15$. Similarly, on the last acquisition trial, the bench means for the Low-Low and Low-High groups are 55.8 and 47.0, respectively, $t(18) = 1.86$, $p > .08$.

Analyses of variance for the acquisition, low-stress transfer, and high-stress transfer conditions are presented, respectively, in Tables 1, 2, and 3. For the acquisition phase the overall means for the Low- and High-stress groups are 80.6 and 96.0, respectively, $F\ (1, 38) = 56.50$, $p < .001$. This result indicates superior performance by the Low-stress group during acquisition trials. (Note: the lower the score, the better the performance.)

Acquisition stress level also appears to affect subsequent transfer performance. In the low-stress transfer situation, the overall means for the Low and High, prior acquisition level, stress groups are 62.9 and 78.0, respectively, $F\ (1, 18) = 21.12$, $p < .001$. Thus, under the low-stress transfer condition, superior performance is demonstrated by the group which previously encountered low stress during acquisition trials.

In the high-stress transfer situation, the overall means for the Low and High, prior acquisition level, stress groups are 94.9 and 82.9, respectively, $F\ (1, 18) = 22.63$, $p < .001$. Hence, under the high-stress transfer condition, superior performance is demonstrated by the group which previously encountered high stress during acquisition trials.

The fact that the Low-Low group is superior to the High-Low group in the

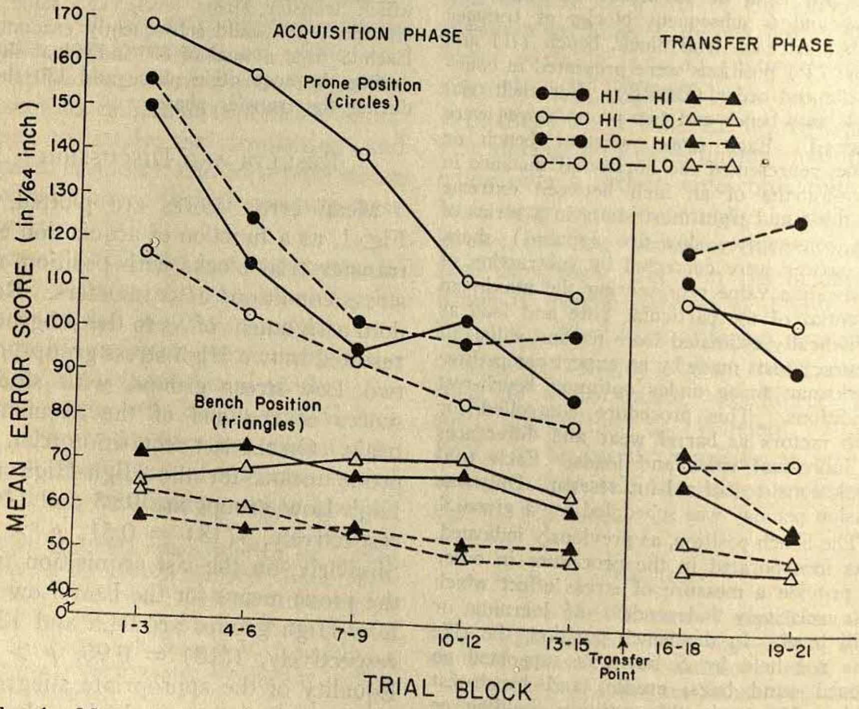


FIG. 1. Mean error scores as a function of acquisition and transfer trial blocks with position and stress conditions as parameters.

TABLE 1

ANALYSIS OF VARIANCE OF ACQUISITION SCORES

Source	*df*	*MS*	*F*
Individuals	38	30,332.17a	
Stress (Low vs. High)	1	1,713,873.70	56.50a***
Trials	14	23,989.70	9.06b***
Positions (Prone vs. Bench)	1	138,122.08	4.53c*
Individuals × Trials	532	2,647.88b	1.02d
Individuals × Positions	38	30,486.62c	11.79d***
Stress × Trials	14	14,747.08	5.57b***
Stress × Positions	1	3,458.72	<1c
Trials × Positions	14	2,972.05	1.15d
Indiv. × Trials × Positions	532	2,585.63d	1.21e
Stress × Trials × Positions	14	2,322.64	<1d
Measurement Error	1200	2,139.00e	
Total	2399		

Note.—Letter following *F* value indicates appropriate error term from *MS* column.
* $p < .05$.
*** $p < .001$.

TABLE 3

ANALYSIS OF VARIANCE OF HIGH-STRESS TRANSFER SCORES

Source	*df*	*MS*	*F*
Individuals	18	13,032.66a	
Stress (Low vs. High)	1	294,921.66	22.63a***
Trials	5	12,892.58	4.80b***
Positions (Prone vs. Bench)	1	17,136.46	1.82c
Individuals × Trials	90	2,685.50b	<1d
Individuals × Positions	18	9,418.82c	2.72d**
Stress × Trials	5	5,591.63	2.08b
Stress × Positions	1	8,737.91	<1c
Trials × Positions	5	1,616.58	<1d
Indiv. × Trials × Positions	90	3,469.47d	1.06e
Stress × Trials × Positions	5	5,555.37	1.60d
Measurement Error	240	3,261.73e	
Total	479		

Note.—Letter following *F* value indicates appropriate error term from *MS* column.
** $p < .01$.
*** $p < .001$.

low-stress transfer situation while the High-High group is superior to the Low-High group in the high-stress transfer situation strongly suggests that a stimulus-generalization factor is involved. However, these transfer results might, in part, reflect the effects of current (transfer) stress level operating directly upon performance and not through a stimulus-generalization mechanism. In order to answer this question it is necessary to have a measure of stress effect which is relatively independent of skill—such as the bench-rest measure included in the present study. In the low-stress transfer situation, the bench means for the Low-Low and High-Low groups are 53.4 and 49.7, respectively, $t(18) = 0.65$, $p > .50$. Similarly, in the high-stress transfer situation, the bench means for the Low-High and High-High groups are 65.8 and 62.3, respectively, $t(18) = 0.62$, $p > .50$. Thus, the bench-rest data suggest that direct effects of current (transfer) stress level on performance can be ruled out as contributing factors to either the superiority of the Low-Low group over the High-Low group or the superiority of the High-High group over the Low-High group.

TABLE 2

ANALYSIS OF VARIANCE OF LOW-STRESS TRANSFER SCORES

Source	*df*	*MS*	*F*
Individuals	18	8,356.52a	
Stress (Low vs. High)	1	176,486.66	21.12a***
Trials	5	738.00	<1b
Positions (Prone vs. Bench)	1	29,047.50	3.48c
Individuals × Trials	90	927.93b	<1d
Individuals × Positions	18	8,340.31c	8.22d***
Stress × Trials	5	361.46	<1b
Stress × Positions	1	45,396.25	5.44c*
Trials × Positions	5	612.79	<1d
Indiv. × Trials × Positions	90	1,014.91d	<1e
Stress × Trials × Positions	5	479.08	<1d
Measurement Error	240	1,088.71e	
Total	479		

Note.—Letter following *F* value indicates appropriate error term from *MS* column.
* $p < .05$.
*** $p <$ 001.

If low stress during acquisition permits the development of a higher level of skill, such a skill differential might explain the superiority of the Low-Low group over the High-Low group but it cannot, simultaneously, account for the superiority of the High-High group over the Low-High group. Hence, irrespective of the direction, or even the existence, of an acquisition skill level differential, stress apparently does affect transfer performance via a stimulus-generalization mechanism.

That stimulus generalization is not the only mechanism by which stress level affects transfer performance is suggested by the analysis summarized in Table 4. In this analysis the difference in column means reflects only the effects of stress level present at the time of measurement, i.e., during the transfer phase. The bench mean is 51.6 for the low-stress transfer group (Low-Low and High-Low groups combined) and 64.05 for the high-stress transfer group (High-High and Low-High groups combined), $F\ (1, 36) = 9.70$, $p < .01$. Thus, the low-stress level apparently produces less direct performance deterioration in transfer than does the high-stress level. Similar results are found in a comparison of acquisition data. The bench means taken over all acquisition trials are 55.10 and 68.05 for the Low- and High-stress groups, respectively, $t(38) = 3.30$, $p < .005$.

The lack of a significant difference in row means, in Table 4, provides evidence supporting the assumption of skill independence of the bench-rest measure. In other words, a differential skill transfer effect, if present, does not show up on the bench measure. Further evidence of this independence is provided by the fact that although performance clearly improves across acquisition trials, $F\ (14, 532) = 9.06$, $p < .001$, the acquisition trial means for the bench measure exhibit only random variation. Specifically, the theory of runs median test of randomness applied to the 15 acquisition trial block means, on the bench measure, for the Low-stress group does not reflect a significant departure from randomness, $p > .05$ ($u = 6$, critical values being 3 and 13). Similar results are obtained in the case of the High-stress group, $p > .05$ ($u = 7$, with critical values being 3 and 13). Thus, under constant stress level over the 15 acquisition trial blocks, the bench measure remains constant except for random variation. As noted above, this constancy cannot be attributed to a general lack of performance improvement. Since the bench measure remains constant over trials and since Tables 1 and 3 both indicate a significant trials effect, one might expect to find a significant trials vs. positions effect. That this interaction was not significant appears to reflect the fact that individuals varied considerably more across positions than across trials, with this difference being reflected in the appropriate error terms.

TABLE 4

Analysis of Variance of Bench Transfer Scores

Source	*df*	*MS*	*F*
Rows (Acquisition Stress)	1	129.6	<1
Columns (Transfer Stress)	1	1,562.5	9.70**
R × C Interaction	1	0.1	<1
Within	36	161.0	
Total	39		

** $p < .01$.

Thus far, results have been presented which suggest that stress affects transfer performance through (*a*) stimulus generalization and (*b*) a direct performance deterioration which is independent of skill. Another question of interest concerns the possibility that stress may affect transfer performance via the level of skill developed during acquisition training. In order to demonstrate that stress affects transfer performance through acquisition skill level, i.e., through the amount of "learning" potentially available at the beginning of transfer, it is necessary to eliminate stimulus generalization and performance effects directly attributable to current (transfer) stress level as alternative explanations. Stimulus generalization can be eliminated as a confounding factor by selecting comparisons for which such a factor is either constant or balanced. For example, the following comparisons might be made: (*a*) Low-Low vs. High-High; (*b*) High-Low vs. Low-High. In both of these comparisons, the role of stimulus generalization is presumably controlled. However, in both of these comparisons the transfer stress level is an uncontrolled factor. The effect of current, or transfer, stress level while not directly controllable in these comparisons is, nevertheless, assessable by means of the bench measure which, as previously noted, appears to be independent of skill.

Considering the first of the above comparisons, the means for the Low-Low and High-High groups on the prone measure during transfer are 72.4 and 103.5, respectively, $t(18) = 1.82$, $p < .05$ (one-tailed test). The superiority of the Low-Low group was predicted on the basis of absence of presumed deteriorative or interfering effects of high stress during both acquisition and transfer conditions. The bench means are 53.4 and 62.3 for the Low-Low and High-High groups, respectively, $t(18) = 1.57$, $p > .05$. Thus, it is reasonable to infer that the superiority of the Low-Low group over the High-High group is due to some factor other than either a stimulus generalization differential or a differential direct effect of current (transfer) stress level. Perhaps the most obvious factor which could account for the observed result is level of skill acquired during acquisition training. Presumably, during acquisition the low-stress condition produced a higher level of skill (learning) than did the high-stress condition.

Additional data consistent with this interpretation are provided by comparing the High-Low and Low-High groups, whose prone means during transfer are 107.3 and 124.0, respectively, $t(18) = 0.98$, $p > .30$ (two-tailed test, no a priori prediction). If, as the preceding comparison suggests, the low-stress condition during acquisition produced a higher level of skill, then the present finding requires, by way of explanation, some factor operating to offset this skill advantage. Since the degree of stimulus generalization is presumably balanced for the two groups, the compensating factor most likely is the effect of current (transfer) stress level. That such an interpretation is tenable is suggested by the fact that the bench means for the High-Low and Low-High groups are 49.7 and 65.8, respectively, $t(18) = 2.84$, $p < .01$ (one-tailed test). The superiority of the High-Low group was predicted on the basis of the assumed skill independence and stress level sensitivity of the bench-rest measure.

REFERENCES

DEESE, J. Skilled performance and conditions of stress. In R. Glaser (Ed.), *Training research and education.* Pittsburgh: Univer. Pittsburgh Press, 1962. Pp. 199–222.

LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 295–317.

WHELEN, T. W. (Ed.) *The ultimate in rifle precision.* Harrisburg, Pa.: Stackpole Company, 1954.

WHELEN, T. W. *Why not load your own!* Washington, D. C.: Combat Forces Press, 1957.

(Early publication received
December 30, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 466–470

DISCRIMINATION SHIFT PERFORMANCE OF CHILDREN AS A FUNCTION OF AGE AND SHIFT PROCEDURE[1]

THOMAS J. TIGHE AND LOUISE S. TIGHE

Dartmouth College

The discrimination shift behavior of children was studied within a $2 \times 2 \times 2$ factorial design which varied age of S (4 yr. vs. 10 yr.), type of shift (reversal, RS, vs. extradimensional, EDS), and state of the irrelevant dimension during shift (constant vs. variable). The 10-yr. olds did not differ in speed of learning RS and EDS under the condition of a constant irrelevant dimension but learned RS faster than EDS when the irrelevant dimension was variable. The 4-yr. olds accomplished EDS more rapidly than RS regardless of the state of the irrelevant dimension. Both types of shift were learned more slowly by the 10-yr. olds when the irrelevant dimension varied than when it was constant, but the shift performance of the 4-yr. olds was not significantly affected by this variable.

There is now a considerable body of evidence indicative of phylogenetic and ontogenetic differences in the relative ease with which organisms accomplish a discrimination reversal (RS) as opposed to an extradimensional shift (EDS, a shift to a previously irrelevant dimension). Briefly, young children (below the age of about 5 yr.) and infrahuman *S*s find an EDS easier than RS, children about 5–7 yr. old do not differ in speed of learning the two types of shift, and mature humans learn an RS faster than an EDS (see Kendler & Kendler, 1962 and T. Tighe & L. Tighe, 1966 for reviews of these data).

Studies with infrahuman *S*s and young children have typically employed a procedure whereby *S* is first given a 2-dimensional discrimination task with one dimension relevant and the other irrelevant and varying within trials. Upon reaching criterion in the initial discrimination, *S* undergoes either an RS or EDS during which the irrelevant dimension is held constant (i.e., varies between trials) in order to minimize partial reinforcement effects for *S*s in an EDS. Eimas (1965) has suggested that the finding of more rapid EDS than RS in these studies may be dependent upon the fact that the irrelevant dimension was held constant during shifts, since this operation serves to reduce the interfering effects of the formerly relevant dimension during EDS but presumably has little or no effect on RS. The implication is that if the irrelevant dimension were to remain variable during shifts, then EDS learning may be slower than RS learning. However, it is clear that relative ease of EDS at the infrahuman level is not dependent upon this aspect of procedure. In a direct assessment of this variable, Kelleher (1956) found that rats accomplished an EDS faster than RS regardless of whether the irrelevant dimension varied or was constant during shift, and the data of Brookshire, War-

---

[1] This investigation was supported by Public Health Service Research Grant MH11088-01, from the National Institute of Mental Health. The authors wish to acknowledge the assistance of David M. Stearns who worked on the project as a National Science Foundation Undergraduate Research Participant under National Science Foundation Grant GY-65.

ren, and Ball (1961) indicate that both rats and chickens learn EDS faster than RS under the condition of a varying irrelevant dimension during original learning and shift. However, it is possible that this procedural variable may contribute significantly to the observed age difference in relative ease of shift at the human level, since comparisons of RS and EDS with adults have not used shift procedures in which the irrelevant dimension is held constant. In a recent experiment addressed to this question, Dickerson (1966) tested preschool children with a variable irrelevant dimension during original learning and shift and found RS to be faster than EDS—a finding in agreement with the data from adults. However, there are aspects of Dickerson's procedure which sharply limit the generalizability of this finding. First, Dickerson did not manipulate the critical procedural variable in relation to shift performance, but simply studied ease of RS vs. EDS under the condition of a variable irrelevant dimension. Thus it is possible that these results are peculiar to the sample of children tested rather than to the shift procedure employed. This point assumes greater significance in view of data indicating that shift behavior may undergo quite abrupt change within a narrow age span (Kendler & Kendler, 1962), and in view of the fact that relative ease of shift varies markedly with degree of perceptual experience with the task stimuli (L. Tighe, 1965). Secondly, Dickerson employed a "special training" procedure whereby a child who did not learn within 40 trials was shown the correct stimuli-reward relations and the incorrect stimuli-no reward relations. In contrast, the comparison studies of interest have required the children to learn the initial discrimination unaided. It is known that the reversal learning of young children can be facilitated by supplementary training (i.e., by overtraining, pretraining, or verbalization procedures), and it is possible that Dickerson's "special training" had a similar effect. Thus, it is clearly desirable to have further information concerning the contribution of differences in dimension presentation procedure to age differences in discrimination shift learning at the human level. The present experiment seeks this information within a factorial design which enables a comparison of the shift performances of children of the same age under a constant vs. under a variable irrelevant dimension procedure and comparison of the performance of children of different ages under the same shift procedure.

METHOD

Subjects.—The *S*s were 64 4-yr. old children (average age 4 yr. 6 mo., range 3 yr. 10 mo. to 5 yr. 2 mo.) and 64 10-yr. old children (average age 10 yr. 6 mo., range 9 yr. 10 mo. to 11 yr. 2 mo.) from the Hanover, New Hampshire area. There were an equal number of boys and girls in each age group.

Procedure.—Each *S* was run individually in a small room alone with *E*. Specific details of apparatus and training procedure can be found in L. Tighe (1965). All *S*s were first trained to discriminate between lightweight metal tumblers varying simultaneously in two values of size (tall, T, 6¼ in., and short, S, 4¾ in.) and two values of brightness (black, B, and white, W, enamel). The stimuli presented in the initial discrimination on any one trial consisted of one of the following two pairs: the TW tumbler vs. the SB tumbler or the TB tumbler vs. the SW tumbler. The *S* was instructed to pick up one of the tumblers on each trial and was told that if his choice were correct he would find a marble under it. The *S* was told that the "game" was to see how soon he could find a marble every time he chose. When *S* made 9 correct out of 10 successive responses, the discrimination shift problem was presented immediately without change of instruction and continued to the same criterion. Half of the *S*s in each age group received an RS and half received an EDS. For half of the *S*s within each shift condition the irrelevant dimension continued to vary within trials while the other half underwent shift with the irrelevant dimension constant (i.e., varying only between trials). Thus for *S*s in the former condition, the same stimulus pairs were presented during the initial discrimination and the shift. The *S*s in the latter condition had to choose during shift between stimuli which varied in only one feature, e.g., an *S* who had "black" as the positive stimulus during shift had to choose between the TB vs. TW tumblers and the SB vs. SW tumblers. Tall, S, B, and W were used equally often as the positive stimulus in each treatment group. Throughout training the stimuli were presented according to a prearranged irregular sequence which equated the frequency of left and right presentations and left and right reinforcement of each stimulus. Each *S* was assigned to groups and positive stimuli on a predetermined random basis. At the beginning of the experiment *S* was shown a large assortment of prizes which included objects of interest to *S*s of both age levels (e.g., comic books, pencils, crayons, bubble gum, M & Ms, jacks set, charms, flutes, etc.). The *S* was also shown a plastic bag containing either 2 quarters (for the 10-yr. olds) or 10 nickels (for the 4-yr. olds). The *S*s were told that they would be given the bag of money when they had finished the game and that they could also pick a prize from the assortment.

RESULTS

The number of trials and number of errors to criterion in the initial discrimination were classified according to the major treatment variables, i.e., age of *S*, type of shift *S* was later to undergo, and state of the irrelevant dimension during shift. Analyses of variance (2 × 2 × 2) of each learning measure established that *S*s in the major treatment groups were from the same population with respect to performance in the initial discrimination (all *F*s < 1.0). Table 1 presents the mean number of trials to criterion during the discrimination shifts for each treatment group. A 2 × 2 × 2 factorial analysis of $\sqrt{X + .5}$ transformed trials to criterion yielded a significant main effect of state of the irrelevant dimension during shift, $F(1, 120) = 21.3$, $p < .001$, a significant interaction between age and type of shift, $F(1, 120) = 25.9$, $p < .001$, and a significant interaction between age and state of the irrelevant dimension during shift, $F(1, 120) = 4.98$, $p < .05$. The *F* values associated with the main effects of age and type of shift and with the remaining interactions were not significant. Comparisons of specific group mean trials to criterion were accomplished by Duncan's multiple-range test. Ten-year olds who underwent shifts with the irrelevant dimension constant did not differ in speed of learning RS and EDS, $p > .10$, but accomplished RS faster than EDS when the irrelevant dimension was variable, $p < .005$. In contrast, 4-yr. olds accomplished EDS more rapidly than RS regardless of whether the irrelevant dimension was constant, $p < .005$, or variable, $p < .05$. For 10 yr. olds learning was slower under the condition of a variable irrelevant dimension than under a constant irrelevant dimension both during RS, $p < .05$, and during EDS, $p < .001$. In the case of the 4-yr. olds, on the other hand, state of the irrelevant dimension

TABLE 1

MEAN NUMBER OF TRIALS TO CRITERION DURING SHIFTS, CLASSIFIED BY AGE OF *S* (4 YR. VS. 10 YR.), STATE OF THE IRRELEVANT DIMENSION (CONSTANT VS. VARIABLE), AND TYPE OF SHIFT (RS VS. EDS)

4-Year Olds				10-Year Olds			
Constant		Variable		Constant		Variable	
RS	EDS	RS	EDS	RS	EDS	RS	EDS
22.9	6.4	38.3	13.3	4.6	6.3	16.2	38.2

did not significantly affect speed of RS or EDS, $p > .10$ in both cases. Analyses of the error measure yielded conclusions identical to those obtained with the trial measure. To summarize the major finding, age differences in relative ease of RS vs. EDS obtained both under a training procedure employing a variable irrelevant dimension and under a training procedure employing a constant irrelevant dimension.

DISCUSSION

These results, together with existing data, indicate that there are reliable age and species differences in discrimination shift learning regardless of the state of the irrelevant dimension during shift. Independent studies comparing ease of shift under the condition of a variable irrelevant dimension have found that infrahuman *S*s and young children find EDS easier than RS, that human adults find RS easier than EDS, and that the maturing child accomplishes these tasks progressively more like the human adult (e.g., cf. Furth & Youniss, 1964; Issacs & Duncan, 1962; Kelleher, 1956; and the data of the present experiment). A similar comparative and developmental trend is indicated for shift behavior under the condition of a constant irrelevant dimension, although data for this condition is lacking at the adult level (e.g., cf. Kelleher, 1956; Kendler & Kendler, 1959; L. Tighe, 1965; and the data of the present experiment). Moreover, overtraining has been shown to enable relative ease of RS in young children (Furth & Youniss, 1964; Tighe & Tighe, 1965) but overtraining does not affect the shift performance of infrahuman *S*s (T. Tighe, 1965; Tighe, Brown, & Youngs, 1965). The results of these comparative and developmental studies, in conjunction with a similar pattern of findings obtained with the "optional shift technique" (Kendler, Kendler, & Learnard, 1962; T. Tighe & L. Tighe, 1966) provide considerable support for the hypothesis that the observed age and species differences in shift behavior reflect fundamental differences in discrimination processes.

The present experiment also provides evidence bearing upon existing theoretical accounts of these data. The fact that the presence of a variable irrelevant dimension significantly retarded the shift performance of the 10-yr. olds but did not significantly affect the 4-yr. olds may be taken to support those explanations of shift behavior which are deduced from the assumption that infrahuman *S*s and young children are likely to discriminate on the basis of the absolute properties of the stimulus objects while the more mature human is likely to discriminate upon the basis of dimensional features (Kendler & Kendler, 1962; L. Tighe & T. Tighe, 1966).

REFERENCES

BROOKSHIRE, K. H., WARREN, J. M., & BALL, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

DICKERSON, D. J. Performance of preschool children on three discrimination shifts. *Psychon. Sci.*, 1966, **4**, 417–418.

EIMAS, P. D. Comment: Comparisons of reversal and nonreversal shifts. *Psychon. Sci.*, 1965, **3**, 445–446.

FURTH, H. G., & YOUNISS, J. Effect of overtraining of three discrimination shifts in children. *J. comp. physiol. Psychol.*, 1964, **57**, 290–293.

ISSACS, I. D., & DUNCAN, C. P. Reversal and nonreversal shift within and between dimensions in concept formation. *J. exp. Psychol.*, 1962, **64**, 580–585.

KELLEHER, R. I. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

KENDLER, T. S., KENDLER, H. H., & LEARNARD, B. Mediated responses to size and brightness as a function of age. *Amer. J. Psychol.*, 1962, **75**, 571–586.

TIGHE, L. S. The effect of perceptual pretraining on reversal and nonreversal shifts. *J. exp. Psychol.*, 1965, **70**, 379–385.

TIGHE, L. S., & TIGHE, T. J. Overtraining and discrimination shift behavior in children. *Psychon. Sci.,* 1965, **2**, 365–366.

TIGHE, L. S., & TIGHE, T. J. Discrimination learning: Two views in historical perspective. *Psychol. Bull.,* 1966, **66**, 353–370.

TIGHE, T. J. Effect of overtraining on reversal and extradimensional shifts. *J. exp. Psychol.,* 1965, **70**, 13–17.

TIGHE, T. J., BROWN, P. L., & YOUNGS, E. A. The effect of overtraining on the shift behavior of albino rats. *Psychon. Sci.,* 1965, **2**, 141–142.

TIGHE, T. J., & TIGHE, L. S. Overtraining and optional shift behavior in rats and children. *J. comp. physiol. Psychol.,* 1966, **62**, 49–54.

(Early publication received March 7, 1967)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 471–475

VARIABLES DETERMINING CONCEPT CONSERVATION IN CHILDREN[1]

IRVING E. SIGEL
Merrill-Palmer Institute

ELI SALTZ AND WILLIAM ROSKIND
Center for the Study of Cognitive Processes, Wayne State University

120 children, 5–8 yr. old, were given problems such as "This father studies and becomes a doctor; is he still a father?" Ability to conserve the concept of father increased significantly with age ($p < .01$). Increasing the distance between the original concept (e.g., father) and its transformed form (e.g., doctor) along both physical and conceptual dimensions significantly decreased the ability of young children to conserve the original concept. These results are discussed in terms of the writer's assumption that children tend to organize information into narrow concepts, and that therefore stimulus alteration leads to concept overdiscrimination.

Piaget (1950) refers to "conservation" as the ability to hold a property invariant in the face of transformations. For example, pouring 4 oz. of water from a tall glass into a flat one does not change the amount of water present, though it does change the appearance of the water. Yet young children, after watching the entire procedure performed before them, will often claim that the flat glass contains less water than the tall one. Piaget interprets such results as indicating that young children cannot conserve the concept of mass.[2] Kooistra (1964) has demonstrated a similar type of problem in the conservation of a less palpable concept, that of "father." Under experimental circumstances, to be described in more detail later, Kooistra told children, "This father studied and became a doctor. Is he still a father?" Many of the younger children insisted that the doctor was no longer a father.

The critical issue in conservation studies is the identification of the variables which influence *S*s' ability to hold a property invariant in the face of transformations. The same issue is central in the more familiar concept-attainment experiments of the type reported by Underwood and Richardson (1956), Richardson (1960), and others. In these latter studies, *S*s are presented with a number of different complex stimuli, all of which have one characteristic in common. The *S*s must learn to identify the crucial characteristic so that they can identify stimuli containing this characteristic as instances of the concept (e.g., milk, chalk, teeth, and snow may be used as instances of the concept *white*). A primary distinction between the concept attainment and the conservation tasks is that in concept attainment *S*s must *discover* the concept, the identification of instances being relatively trivial once discovery is made. In conservation, *S*s are told the concept and the identification of instances is then tested.

Several hypotheses have been offered

---

[1] The present research was supported by National Science Foundation Grant GB 2538.

[2] This problem in conservation is the basis of the old riddle, "Which weighs more, a pound of feathers or a pound of lead?"

for the relative inability of children to conserve concepts in the type of tasks outlined above. Bruner (1964) and Braine and Shanks (1965) suggest that the difficulties encountered by children in conserving physical concepts (e.g., mass, length) are essentially due to perceptual factors. Bruner suggested that children are misled by the perceptual elements. When children do not actually observe the water being poured from a tall, narrow glass to a flat glass, but are merely told of this operation, they are more likely to say that the amount of water was the same in the narrow glass as in the flat one (Bruner, 1964). Braine and Shanks, on the other hand, feel that young children are not necessarily deceived in such problems. Instead, they interpret *E*'s question as asking which glass *looks* as though it contains more water.

Such interpretations appear to be restricted, since they do not cover the more abstract types of conservation problem such as Kooistra's. Saltz and Sigel (1967) have suggested that children tend to possess relatively narrow concepts containing few attributes. Therefore, if an instance occurs containing attributes markedly different from the standard, the child interprets it as a new concept. Consequently, younger children will display less tolerance for differences between a standard and an instance of a concept than older children.

In examining this proposition in the present experiment, problems similar to those employed by Kooistra were used. Variations along several different dimensions were manipulated: (*a*) *Time*. If a standard and an instance are both presented as existing in the same time, there is less difference between them than if they are presented as in different times. (E.g., "This father is a doctor" presents both standard and instance in the same time period. "This father studied and became a doctor" involves a transition in time between a present father and a future doctor.) (*b*) *Number of Transitions*. The greater the number of transitions through which an instance goes, the greater the difference between the instance and the standard. (E.g., A father who is a doctor undergoes one transition. A father who is a doctor and is such a good doctor that he is president of the hospital, undergoes two transitions.) (*c*) *Physical Representation*. If one group of pictures is labeled "fathers," and a different group is labeled "doctors," *S* is placed in a situation in which there is an objective difference between fathers and doctors (i.e., the two sets are distinct). If the picture of an instance of "father" is moved from the "father" set to the "doctor" set, this represents a greater difference between the instance and the set of "fathers" than if *S* is merely told, "This father is a doctor."

In essence, irrelevant cues of a temporal, transformational, or representational nature were introduced in each situation. Consistent with our earlier findings on concept overdiscrimination, an inverse relationship between conservation and irrelevant contextual characteristics was expected.

METHOD

Subjects.—The *S*s were 120 children from the summer camp of the City and Country School, Bloomfield Hills, Michigan. The *S*s were 5, 6, 7, or 8 yr. of age, with 30 *S*s in each age group. The 30 *S*s in each age group were randomly assigned to the five experimental conditions with the restriction that each age was represented by 6 *S*s in each condition.

Materials.—Each *S* was presented three conservation problems. The core problems for all groups involved: (*a*) a father who is a doctor; is he still a father? (b) a mother who is a schoolteacher; is she still a mother? (*c*) a brother (or sister, depending on the

TABLE 1

SUMMARY OF THE FIVE EXPERIMENTAL GROUPS

Group	Time Relationship	Number of Transitions	Physical Representation
1	is	1	yes
2	studied and became	1	yes
3	is	2	yes
4	studied and became	2	yes
5	studied and became	1	no

Note.—Each group consisted of 24 *S*s, 6 *S*s at each of the following age levels: 5, 6, 7, and 8 yr.

sex of *S*) who is a schoolboy; is he still a brother?

Problems involving an alteration in time substituted "studied and became" in place of "is" for each of the above core problems. All the core problems above involve a single transition (e.g., father-doctor). The two transition problems involved (*a*) father-doctor-president of the hospital; (*b*) mother-teacher-principal of the school; (*c*) brother-schoolboy-teacher's helper. In each problem, the second transition was kept in the same time mode as the first (e.g., is such a good doctor that he is president . . . vs. became such a good doctor that he was made president . . .).

Eight cutout representations of men were used in conjunction with Problem *a;* of women in conjunction with Problem *b;* of boys (or girls) in conjunction with Problem *c*.

Procedure.—The following is an example of the instructions and first question used for Experimental Group 1.

> I am going to show you some dolls and ask you some questions after I tell you about the dolls. Here are four dolls who are fathers and here are four dolls who are doctors. (All dolls were placed in a row in front of *S*, a large space separating the fathers from the doctors.) Now, which are the fathers and which are the doctors?
>
> One of the fathers gets up early in the morning and because he is a doctor, he joins the doctors group at work. (At this point one of the dolls from the fathers group was moved to the group of dolls which had been identified as doctors.) Is he still a father?

The five experimental groups are summarized in Table 1. For any *S*, all three problems involved the same time relationships, number of transitions, and presence or absence of physical representation.

RESULTS

There was a marked tendency for the scores to be distributed bimodally. Most *S*s either were correct on all three problems, or failed all three. Very few scores of 1 or 2 were observed. Since the effects of such distributions on analysis of variance are not completely understood, a multidimensional chi-square analysis (see Winer, 1962) was performed. Table 2 summarizes the number of *S*s solving all three problems in each condition.

Of the 96 *S*s in the $2 \times 2 \times 3$ analysis, 52 *S*s solved all three problems correctly. Consequently, the chi-square analysis scored three correct as "plus" and less than three correct as "minus."

Considering first the effect of manipulating time, of the 48 *S*s in the unaltered time condition (e.g., father *is* a doctor), 69% solved all three problems. Of the 48 *S*s in the altered time condition (e.g., father *studied and became*), only 39% solved all three problems, $\chi^2(1) = 8.2237$, $p < .01$.

Percentages of *S*s solving all three problems at the 5, 6, 7, and 8 yr.-old levels, respectively, were 37%, 50%, 50%, and 79%, with 24 *S*s at each age level, $\chi^2(3) = 9.062$, $p < .05$.

TABLE 2

NUMBER OF *S*S IN EACH GROUP SHOWING CONCEPT CONSERVATION ON ALL THREE PROBLEMS

Age	Experimental Conditions				
	1	2	3	4	5
5	2	1	5	1	4
6	4	2	4	2	3
7	5	3	3	1	4
8	5	4	5	5	6

Number of transitions (e.g., father-doctor vs. father-doctor-president of hospital) had no effect on conservation, with identical percentages of *S*s, 54%, solving both the one and two transition problems. In addition, none of the chi-squares for the interaction terms approached significance.

An evaluation of the effect of physical representation (e.g., a doll called "father" is taken from a group of other "father" dolls and placed in a group of dolls called "doctors") was performed in a separate 2 × 4 factorial design which compared the conservation performance of Experimental Group 2 with Experimental Group 5 of Table 1. These groups are identical except for the presence of physical representation in Group 2 and its absence in Group 5.

Of the 24 *S*s for whom physical representation was present, only 42% solved all three problems; of the 24 *S*s without physical representation, 71% solved the problems, χ^2 (1) = 4.1447, $p < .05$. To evaluate the effect of age on these problems, the 5 and 6 yr. olds were combined into one group of 24 *S*s and the 7 and 8 yr. olds into a second group of 24 *S*s. Of the older group, 71% solved the problem, while 42% of the younger group solved it. By coincidence, these percentages are identical with those obtained for physical representation, and so the chi-square is also identical (4.1447) and significant. The interaction chi-square between age and physical representation failed to approach significance.

DISCUSSION

In the present study, presence of physical representations resulted in poorer concept conservation than absence of such representations (e.g., moving a doll from a group called "fathers" to a group called "doctors" resulted in more children saying the doctor was no longer a father). These results are consistent with Bruner's position that failure of conservation is due to competing perceptual cues. They are consistent also with the Braine and Shank's position that children respond as though the question asked by *E* was, "Does it look as though the doctor is a father"; and they are consistent with the present writers' position that children have narrow concepts which are disrupted by stimulus alteration.

However, the results arising from manipulation of the verbalized time dimension appear to be inconsistent with both the Bruner and the Braine and Shanks positions. Children conserved significantly better when verbalized time relations were not altered (e.g., this father *is* a doctor) than when time was altered (e.g., this father *studied and became* a doctor). Since such verbalizations concerning time are equivalent in degree of *perceptual* alteration, the significant difference cannot be attributed to perceptual factors.

In the present study, manipulation of the perceptual irrelevant cues of physical representation and the verbal irrelevant cues of the time dimension had approximately equal effects on conservation. This can be seen by considering as a base line the condition involving both the irrelevant physical representation and the altered verbalization concerning time (i.e., Groups 2 and 4). In this condition, only 39% of the *S*s conserved on all three problems. When the irrelevant cues of physical representation are removed but the altered time factor is retained (Group 5), conservation is attained by 71% of the *S*s. Similarly, when the altered time factor is eliminated but physical representation is retained (Groups 1 and 3), conservation is attained by 69% of *S*s. It appears, then, that perceptual cues do not have a special status in this type of a problem. Irrelevant conceptual cues may be as effective as irrelevant perceptual cues in impeding conservation.

Of the three stimulus dimensions manipulated in the present study, only *number of transitions* failed to have a significant effect on concept conservation. Lack of differences in the present type of study leaves us in a particularly ambiguous position. The failure to find an effect due to manipulating number of transitions could simply be the result of comparing points which were too close to one another along the dimension. On the other hand, there may be qualitative characteristics of certain dimensions which make them more susceptible to the effect. The present data provide no basis for choosing between these alternatives.

Failure to obtain an effect due to manipulating number of transitions does appear to eliminate one type of variable as a factor in the significant effects obtained with the other dimensions. The two-transition condition involved more words, and a greater possible taxing of memory for the young *S*s, than the one-transition condition. The fact that there is no difference in concept conservation between one- and two-transition conditions suggests that small differences in amount of verbiage, per se, is not an important factor for the phenomenon.

Finally, the present results appear to raise questions about a "stage" theory of conservation. The evidence is contrary to the position that, at a certain stage of cognitive development, ability to conserve certain types of concepts, (e.g., the concept of weight, or of father) emerge. Instead, with increased age we find a steady increase in the degree of irrelevant stimulation that can be tolerated in identifying a given concept.

REFERENCES

BRAINE, M. D. S., & SHANKS, B. L. The development of conservation of size. *J. verbal Learn. verbal Behav.*, 1965, **4**, 227–242.

BRUNER, J. S. The course of cognitive growth. *Amer. Psychologist,* 1964, **19**, 1–15.

KOOISTRA, W. H. Developmental trends in the attainment of conservation, transitivity, and relativism in the thinking of children: a replication and extension of Piaget's ontogenic formulations. Unpublished doctoral dissertation, Wayne State University, 1964.

PIAGET, J. *The psychology of intelligence.* London: Routledge and Kegan Paul, 1950.

RICHARDSON, J. Association among stimuli and the learning of verbal concept lists. *J. exp. Psychol.*, 1960, **60**, 290–298.

SALTZ, E., & SIGEL, I. E. Concept overdiscrimination in children. *J. exp. Psychol.*, 1967, **73**, 1–8.

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 25, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 476–484

SOME TEMPORAL CHARACTERISTICS OF VISUAL PATTERN PERCEPTION[1]

CHARLES W. ERIKSEN AND JAMES F. COLLINS

University of Illinois

Visual stimuli were constructed so that any given stimulus by itself appeared to be a random collection of dots. However, when 2 corresponding stimuli were superimposed by means of a 2-field tachistoscope, a 3-letter nonsense syllable was perceived. Temporal organization in perception was studied in Experiment I by varying the interval between the presentation of the 2 corresponding patterns over 300 msec. Identification accuracy of syllables was a decreasing function of interstimulus interval over a range in excess of 100 msec. Experiment II used unequal energy levels for the 2 corresponding patterns and also varied the sequence of occurrence of the high and low energy members of corresponding sets. The relevance of such concepts as perceptual memory, afterimages, and psychological moments to the data was considered. It was suggested that mechanisms in the visual system such as the "off" response that detect the termination of a stimulus may be responsible for inhibition of integration over time.

The visual perceptual system under certain circumstances is unable to resolve short time differences. Bloch's law relates to the fact that the perceived brightness of the stimulus is dependent not only upon its luminance but its duration. Recent work indicates that this critical duration over which time intensity reciprocity holds is valid not only for perceived brightness but also for form identification. In the latter case the values of the critical duration are somewhat longer than those found for luminance summation (Kahneman & Norman, 1964). These findings would imply that the visual perceptual system sums or integrates energy over the critical duration prior to the occurrence of the perception of a form or of a brightness magnitude.

Experiments upon perception of simultaneity and of perceptual rate have indicated a perceptual time unit generally in the region of 50–100 msec. (see White, 1963, for an excellent review of these studies). These and the reciprocity studies have proven difficult to conceptualize. Attempts to deal with these temporal factors in perception have led to the conception of a psychological moment (Pièron, 1952; Stroud, 1956; White, 1963; White & Cheatham, 1959) that has many of the characteristics that have been ascribed to psychological time units by the philosophers Bergson (1913) and James (1890). Renewed interest in a psychological moment has come from research on backward and forward masking in visual perception. At least certain masking effects can be interpreted as a temporal summation of test and masking stimuli into a composite perception (Eriksen, 1966; Eriksen & Collins, 1965), and Boynton (1961) has suggested an explanation for tem-

---

[1] This investigation was supported by Public Health Service Research Grant MH-1206 and a Public Health Service Research Career Program Award K6-MH-22,014. This paper was presented in part at the Symposium on Temporal Factors in Vision and Visual Perception, June 1966, sponsored jointly by National Science Foundation and the Center for Visual Science, University of Rochester.

poral summation in vision and of masking phenomena in terms of a psychological moment.

Studies of temporal factors in visual perception, including the masking experiments, typically have dealt with brightness and acuity tasks and less frequently with form identification. However, Kahneman (in press) has suggested that the latter task as generally employed is essentially also an acuity task. In the present experiments we have employed a technique of stimulation that permits the study of the temporal development of organizational or integrational components in pattern perception. The first two lines in Fig. 1 show what appear to be two random dot patterns. When these dot patterns are superimposed they yield the perception of the nonsense syllable VOH. Each of the dot patterns, hereafter designated as a stimulus half, can be placed in the separate field of a tachistoscope and when the halves are properly aligned and exposed simultaneously at an appropriate luminance and duration, they yield the perception of the nonsense syllable. By introduction of systematic delays between presentation of the halves the temporal distribution of an organizational component in perception can be studied.

This task has several advantages. Not only does it permit the study of the temporal course of organizational components in perception similar to those that underlie perception of imbedded and street-gestalt figures, but it permits the study of a possible perceptual memory as suggested by Averbach and Coriell (1961) and Sperling (1963). As the reader can verify for himself, the two dot patterns are not such that they can be cognitized and/or remembered in such a manner as to yield the contained nonsense syllable. The nonsense syllable would seem to be capable of being perceived only if the two halves are perceived as psychologically simultaneous or the perceptual trace of the first half is still present when the second stimulus half occurs.

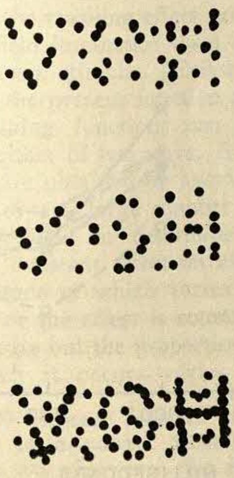


FIG. 1. The upper two dot patterns when superimposed result in the bottom stimulus pattern in which the nonsense syllable VOH can be read.

EXPERIMENT I

Subjects.—Five male graduate students with normal or corrected to normal vision served as paid *O*s.

Procedure.—A Scientific Prototype Model GA three-field tachistoscope was used for stimulus presentation. The "complete" stimuli consisted of 20 three-letter nonsense syllables. The first letter was always a T, V, or X, the second A, O, or U, and the third H, M, or W. As seen from the sample stimulus in Fig. 1, the letters were composed of dots. The dots making up the forms of the letters were divided over two cards so that the dots contained on either card alone yielded minimum information as to the nature of the nonsense syllable. To reduce cues further and to minimize the possibility that the nonsense syllable could be guessed from the dots on only one stimulus card, slightly smaller camouflaging dots were distributed over each card. This technique yielded stimulus halves from either of which alone it was impossible to perceive the imbedded nonsense syllable. These stimulus cards were originally made up of India ink dots on white cardboard. Before use they were photographed to yield high-contrast negatives. The photographic negatives of the stimulus halves were presented in separate fields of the tachistoscope which had been carefully aligned so as to superimpose properly the corresponding halves thus permitting the nonsense syllable to be perceived.

As presented in the tachistoscope the stimulus halves subtended 2° of visual angle in length and 1° in height. They were backlighted with a luminance of 5 mL. in each field. The third tachistoscopic field was dark but contained a faintly luminous cross .3° of angle and positioned .5° below the center letter of the nonsense syllable.

The *S*s were instructed to fixate the cross and to trigger a stimulus presentation when the cross appeared sharp and clear. An experimental trial consisted of the presentation of both stimulus halves either simultaneously or at an experimentally varied interstimulus interval (ISI). The *S*s had available in their dimly illuminated booth a list of the 20 possible nonsense syllables from which they selected their response following each trial. On the average, a 5-sec. interval occurred between trials. Prior to beginning the experimental sessions each *S* served two practice sessions during which the corresponding stimulus halves were presented simultaneously for a duration of 6 msec. All *S*s obtained an accuracy level of at least 85% but less than 100% for identification of the nonsense syllables during the second practice session (average 93%). Following the practice sessions *S*s were run for five additional experimental sessions in which the stimulus halves were separated in time with ISI values of 0, 25, 50, 75, and 100 msec. between the offset of the first half and onset of the second stimulus half. In each session a block of 20 trials was run for each delay. The delays were counterbalanced over sessions and *S*s to cancel any order effects.

Following completion of the five experimental sessions, examination of the data failed to reveal a definitive asymptote in identification as the ISI was increased. To determine whether some temporal summation was still occurring at the 100-msec. ISI, *S*s returned for three additional experimental sessions during each of which they were run on ISIs of 100, 300, and 500 msec.

RESULTS AND DISCUSSION

In Fig. 2 mean percentage of correct nonsense-syllable identification is plotted as a function of ISI between the offset and the onset of corresponding halves. The average function is

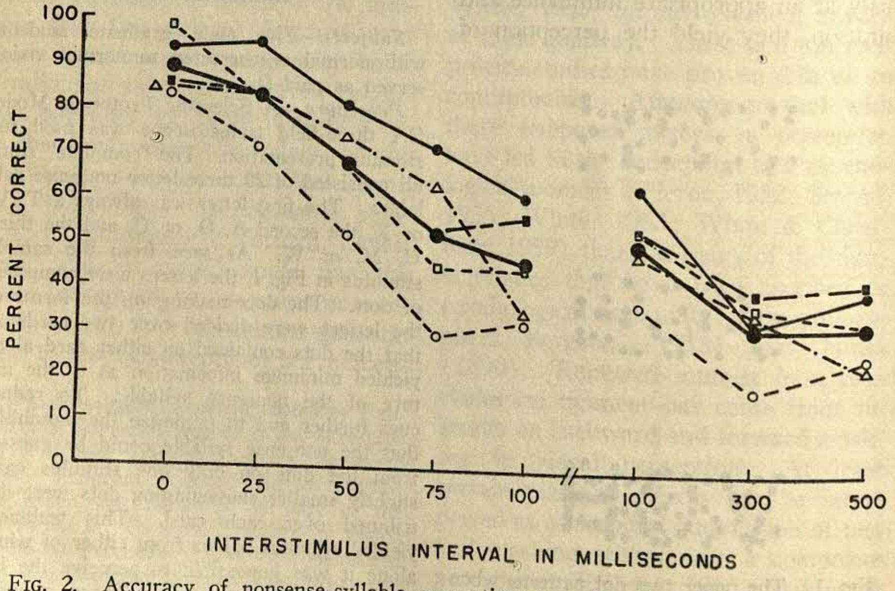


FIG. 2. Accuracy of nonsense-syllable perception as a function of ISI between the corresponding stimulus halves. (The heavy lined function represents the average and the lighter lined functions are for the individual *S*s.)

shown by a heavy line and the data for the five *S*s is shown in the lighter lined functions. The three additional sessions at ISIs of 100, 300, and 500 msec. that *S*s served following the counterbalanced experimental sessions are plotted separately in Fig. 2. For three of the five *S*s there is little if any drop in performance when the halves are presented over the range of simultaneity to an ISI of 25 msec. For all *S*s performance decreases at an accelerated rate between ISI values of 25 and 75 msec. It appears clear that an asymptote in identification is approached at somewhere between an ISI of 100 and 300 msec.[2]

Phenomenal reports obtained from *S*s indicated that nearly all were detecting the double stimulation or the separate halves by ISI values of 75 msec. At this separation *S*s reported a twinkling effect in the stimulation due to apparent movement of the dots toward adjacent dots. By 100-msec. ISI most *S*s could perceive the two halves as separate stimulations.

The decreasing identification performance as a function of increasing ISIs is similar to the masking effect produced by a second-field luminance upon identification of a form stimulus (Eriksen, 1966). However, the present function as well as many masking functions can be interpreted in either of two ways. Since these functions are obtained by averaging performance over a large number of trials, it is impossible to determine whether there is a constant effect on every trial, the magnitude of which varies inversely with ISI, or the effect is constant whenever it occurs but the proportion of trials upon which it occurs varies inversely with the ISI. For the present data the first alternative might be phrased in terms of a decaying trace of the first stimulus half. The longer the ISI the more the trace or "perceptual" memory of the first half has decayed. When the second half is presented *S* must put together or organize the decayed or less intense trace of the first half with the brighter second-half pattern in order to perceive the nonsense syllable. Thus at any ISI over the range where some integration occurs the difficulty of the organizational task is essentially constant for that given ISI on each trial. This would assume that the integration of the nonsense syllable from the two dot patterns is impaired if the halves are of unequal brightness.

In terms of the second alternative the obtained function is generated by an underlying process by which both halves are perceived simultaneously on some proportion of the trials. In this case performance is unimpaired. But on the remaining trials they are perceived as two separate dot patterns and integration is not possible. This second possibility would be commensurate with a concept of a psychological moment (Boynton, 1961; Murphree, 1954; Stroud, 1956; White, 1963).

If one postulates that experience and/or perception occurs in chunks or quanta of somewhere in the neighborhood of 50–100 msec. duration, the present data would be interpreted in the following manner. At short ISI values the probability that both stimulus halves would occur in the same psychological moment and therefore be perceived as though presented simultaneously is quite high. However, as the ISI value increases the probability decreases that both halves would enter the same psychological moment. By the time the separation of the two halves is greater than the duration of the psychological moment, identification performance would have reached an asymptote since under these circumstances the probability is zero that they could enter the same moment.

Other explanations are of course possible and the present data do not provide even a choice between the decaying perceptual trace or memory explanation and the psychological moment concept. Experiment II was designed to test the appropriateness of a decay interpretation and also the possibility that the integration over time effects in the above study were due to subtle afterimages. This latter possibility is operationally difficult to distinguish or disentangle from a perceptual memory or a decay process and in the present paper no attempt is made to draw a distinction.

---

[2] With 20 nonsense syllables the a priori probability of a correct identification by guessing is .05. However, the inability to completely eliminate the partial cues in the separate stimulus halves required the determination of an empirical guessing accuracy level.

EXPERIMENT II

Design.—One of the stimulus halves is given a longer duration than its corresponding half. The energy of the longer duration half is thus greater and in keeping with Bloch's law would be expected to constitute a brighter or more intense stimulation. In the present experiment one of the stimulus halves was always presented at a 25-msec. duration at 5-mL. luminance while the corresponding stimulus half was presented at either 50-, 100-, or 150-msec. duration at 5-mL. luminance. A second variable was whether the longer duration stimulus half preceded or followed the corresponding 25-msec. duration half. Each combination of these two variables was presented at three ISIs: immediately successive (less than 1 msec. apart); 20 msec.; and concurrently. In the concurrent condition, due to the inequality of the duration of the stimulus halves, they were actually concurrent only for the 25-msec. duration of the shorter half. The order variable was manipulated by having the corresponding halves have a concurrent onset or a concurrent offset. Thus, under the concurrent condition when the long duration half occurred first its onset preceded the short duration half by the amount of the time difference in their duration but they shared concurrently the last 25-msec. of the duration of the longer half. For the reverse order both halves had a concurrent onset but after 25 msec. the short-duration half terminated and the long-duration half continued for the length of its remaining duration interval. In addition to these conditions control observations were systematically obtained for the conditions where both of the corresponding halves had a 25-msec. duration which was presented either concurrently, immediately successive, or at 20-msec. ISI.

This design should permit answers to the following questions raised by the preceding experiment. First of all, a decaying trace interpretation of the above data requires that integration of the nonsense syllable from the corresponding stimulus halves is more difficult if the halves are of unequal energy than if the halves are of comparable energy. In the present experiment this assumption would lead to the expectation that identification performance should decrease as the duration of one of the stimulus halves varies from 25 msec. to 150 msec. and the other half remains constant at 25-msec. duration. The decaying trace explanation is tested most directly by the order variable. It seems reasonable to expect that the absolute energy in a decaying perceptual memory (or the clarity of an afterimage) at some succeeding moment in time would be greater for a high-energy stimulation than for a low-energy stimulation. For example, at an ISI of 20 msec., identification of the nonsense syllable should be easier if the first stimulus half had a longer duration since there would be a more persistent trace to integrate with the following short-duration half. The reverse order should lead to much worse performance, not only due to less persistence from the low-energy half having occurred first, but also because under these circumstances there would be a greater absolute difference in energy of the corresponding halves. Further, the decaying perceptual memory explanation would expect an interaction between long-half duration and ISI since as ISI increases, high-identification performance could be maintained by increasing the energy in the first half presentation and therefore the amount of persistence of the memory trace.

Subjects.—Six undergraduate students (four males) served as paid volunteers. All had normal or corrected to normal visual acuity and none had served in Exp. I.

Apparatus and stimuli.—Apparatus and stimuli were essentially the same as those employed in Exp. I with these exceptions. Previously the dot patterns of the stimulus halves had been backlighted thus presenting a view of illuminated dots against a black ground. In this experiment the figure-ground relationship was reversed. The stimuli were made using India ink on white cards and were frontlighted in the tachistoscope at a luminance of 5 mL. To prevent luminance summation—contrast reduction effects that would vary as a function of ISI (Eriksen, 1966), the adapting field which contained a small black fixation cross also had 5-mL. luminance. To maintain contrast reduction due to luminance summation as a constant overall ISI and other experimental conditions, the adapting field was left on continuously except where the conditions required that both stimulus halves be on concurrently. Then the adapting field was off while the stimulus halves were concurrently presented but when either half terminated the adapting field came on instantaneously. As constructed on the white cardboards the black dots had a contrast of 95% but due to "veiling glare" arising from superimposition of adapting and stimulus fields or of the two stimulus fields at the eye the effective contrast was 47.5% (Kahneman, 1966).

Procedure.—The procedure was also essentially the same as in Exp. I. Each *S* served two practice sessions in which the 20 sets of corresponding stimulus halves were presented simultaneously for 25 msec. All *S*s who were subsequently used in experimental sessions attained an accuracy level of at least 85, but less than 100% during the last practice session. Several potential *S*s had to be dismissed because of their inability to achieve this criterion. Following the two practice sessions each *S* served for 12 experimental sessions. During these sessions each *S* had 40 trials under each combination of the 18 experimental conditions (three durations of the long stimulus half, 50, 100, and 150 msec.; two orders, long half first or second; and three ISIs, concurrent, 0- and 20-msec. delay between halves). In addition, 80 trials were run under the control conditions at each of the three ISI values where both stimulus halves were 25-msec. duration.

RESULTS

The number of correct nonsense-syllable identifications was analyzed in a four-way classification analysis of variance (three durations of the long-stimulus half, order of occurrence of long- and short-duration stimulus halves, ISI, and *S*s). All main effects and interactions were significant beyond the .01 level except the main effect due to order of occurrence of long and short half, and the following interactions: Long-half duration with *S*s; ISI, long-half duration and *S*s; ISI, order and *S*s; and long-half duration, order and *S*s. The control condition for the equal duration halves at the three ISI values were analyzed separately and a significant effect ($p < .001$) was obtained for the effect of ISI.

In Fig. 3 average percentage of correct nonsense-syllable identification is shown as a function of the duration of the long-stimulus half. The parameters in Fig. 3 are the three ISI values: concurrent, 0, and 20 msec. For each of these ISI values the effect of order (whether the long duration-stimulus half preceded or followed the short-duration half) is also shown. The points at the extreme left of Fig. 3 are for the control condition where the corresponding stimulus halves each had a duration of 25 msec. and were ob-

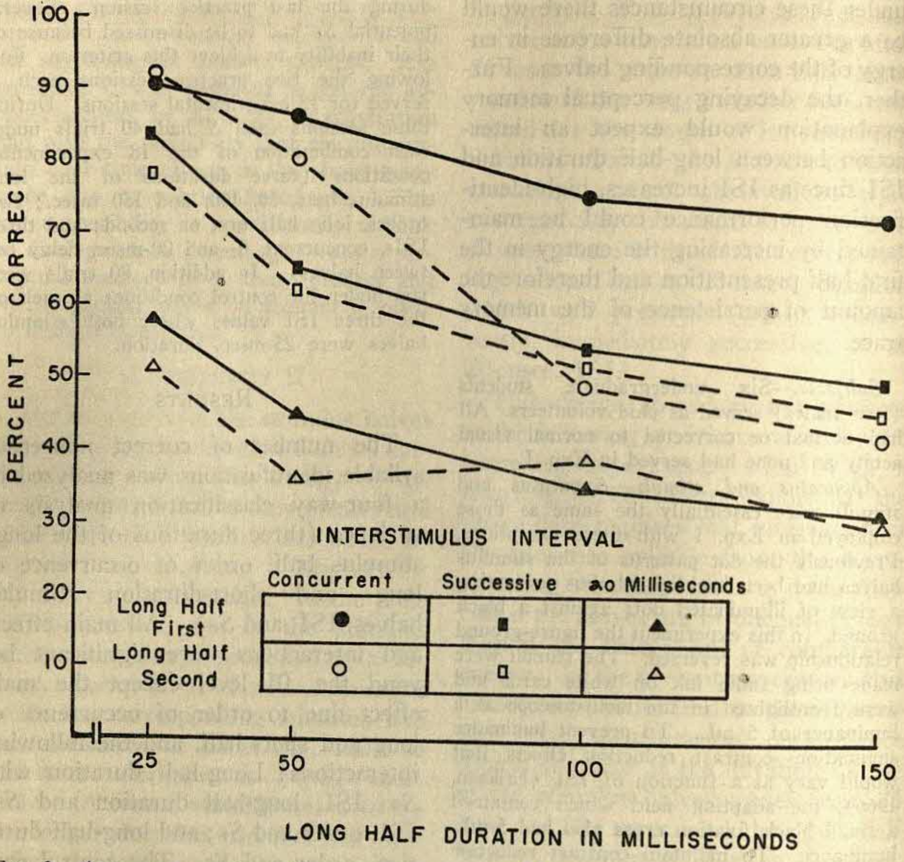


FIG. 3. Average percentage of correct nonsense-syllable perception as a function of the duration of long-stimulus half, ISI, and sequence of long- and short-stimulus halves.

tained under each of the three ISI values.

The data of Fig. 3 may be summarized as showing that the ability to organize the nonsense syllables from the separate stimulus halves is: (*a*) greatest when the stimulus halves are concurrent and decreases as the halves are separated in time; (*b*) highest when the stimulus halves are of equal energy (duration) and decreases in a negatively accelerated function as the mismatch in energy increases; and (*c*) indifferent to the order of occurrence of unequal energy-stimulus halves except when the occurrence of the stimulus halves overlap in time (concurrent). In this latter instance performance is best if the unequal duration stimulus halves share a common offset rather than a common onset, and the greater the inequality in durations the greater is this effect.

DISCUSSION

The above data support the assumption that inequality in energy between stimulus halves reduces integration or organization of the imbedded nonsense syllable. This assumption is basic to a decay explanation. However, a simple decay explanation is countermanded by the fact that, (*a*) when the stimulus halves are not concurrent, there is no performance gain when the high-energy stimulus half occurs first rather than after the low-energy half, and (*b*) there is a progressive decrease in performance when the occurrences of the stimulus halves are varied from concurrent to successive to 20-msec. ISI for all duration or energy inequalities.

It seems reasonable that if a perceptual memory process exists, the greater the intensity or energy of the stimulation, the greater the intensity of the trace or memory at a succeeding point in time. Thus in the present study when a long duration half is presented first, there should be a more intense memory trace to integrate with the short-duration half when it occurs later in time. The reverse order should lead to inferior integration, not only because the short-duration trace is less intense but also because the energy inequality between the stimulus halves will also be greater. Since integration is best when the halves are of equal energy it also follows that an interaction of first-half duration with ISI should occur. This would represent the instances where the decay of the more intense first-half stimulus now matches more closely the intensity of the succeeding short-duration half.

As was seen, there were no indications in the above data to support these deductions. The only case where order of occurrence of long- and short-duration halves made a difference was when the halves were on concurrently for the duration of the short half. For example, if the long-stimulus half had a duration of 100 msec. and the short half 25 msec., perception of the nonsense syllable was more accurate if the 100-msec. stimulus half was turned on first and the short-duration half turned on 75 msec. after the onset of the long-stimulus half. If both stimulus halves were turned on simultaneously and 25 msec. later the short-duration half terminated but the long half continued for an additional 75 msec., perception of the nonsense syllable was much inferior.

This finding is in itself puzzling. It does not fit a memory decay process since both halves always share a common 25-msec. duration. Performance under either order is poorer than if both halves had been presented for only 25 msec. concurrently. The inequality of the energy in the stimulus halves would seem to account for the decreased performance except that this inequality should be the same for either order of the halves occurrence. When the halves share a common onset there is a 25-msec. period when they are of the same energy but it is apparent that perceptual integration does not occur that rapidly.

The interaction also is not readily interpretable in terms of a simple conception of a psychological moment. Since both halves always share a common 25-msec. duration, at least this much would be included in the same psychological moment, and perception of the imbedded nonsense syllable would be expected to be quite accurate. Some deterioration in performance could be attributed to the variability of the beginning of the psychological moment such that on some trials more than 25 msec. of the energy of the long stimulus-half duration might be included in the same moment with the 25-msec. duration half. This would lead to some energy mismatch between the two halves with a corresponding decrease in identification performance. However, the operation of such a process would be independent of the order of occurrence of the long- and short-duration halves and would not be commensurate with the obtained interaction.[8]

Those who have discussed the concept of the psychological moment (Boynton, 1961; Murphree, 1954; Stroud, 1956; White, 1963) have not been sufficiently specific in describing its characteristics. To determine whether the concept has theoretical value in handling phenomena such as produced in the present experiments we need greater explication. Does an arriving stimulation enter an ongoing psychological moment or is it stored and then represented in the succeeding moment? In the latter case when it's stored does any deterioration occur in its trace or intensity during the period of storage? Does the duration of the moment vary with intensity of stimulation? Specifications such as these are essential if the moment concept is to be applied to the interpretation of the present data. Failure of the theorists to deal with these details of the nature of the concept is undoubtedly attributable to the paucity of experimental data on temporal factors in perception.

---

[8] Fred Attneave (personal communication) has suggested that the interaction of order of occurrence with the long and short-duration halves might be attributable to apparent movement. When the stimulus halves have a common onset but terminate 75–125 msec. apart, the short-duration half is assimilated into the long-duration half via apparent movement. This process would account for the unidirectional effect since if the halves have an asynchronous onset but common termination the conditions for apparent movement would not obtain.

The results of the present experiments are most likely a product of several different processes in the visual perceptual system. It seems apparent from the present data alone that perception lags at least 25 msec. or more behind the onset of stimulation in physical time. Differential latencies in transmission of stimulation from sense organ to higher brain centers may well be one factor in the present results. Perhaps most important is the consideration that a system which has the summative and persisting characteristics of visual perception would seem to require other mechanisms to detect discontinuity or interruptions in stimulation. The "off" response recorded at the level of the retina (Granit, 1947) would seem to be one such mechanism. It is possible that this mechanism or similar mechanisms that detect discontinuity in stimulation have an inhibiting effect on the summative or the integrative process underlying performance in the present experiments. This would be commensurate with the finding that any discontinuity or interruption of stimulation was found to have a deleterious effect on perceptual integration of the imbedded nonsense syllables.

REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell Sys. tech. J.*, 1961, **40**, 309–328.

BERGSON, H. *Creative evolution.* New York: Holt, 1913.

BOYNTON, R. M. Temporal factors in vision. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 739–756.

ERIKSEN, C. W. Temporal luminance summation effects in backward and forward masking. *Percept. Psychophys.*, 1966, **1**, 87–92.

ERIKSEN, C. W., & COLLINS, J. F. A reinterpretation of one form of backward and forward masking in visual perception. *J. exp. Psychol.*, 1965, **70**, 343–351.

GRANIT, R. *Sensory mechanisms of the retina.* New York: Oxford Univer. Press, 1947.

JAMES, W. *The principles of psychology.* Vol. 1. New York: Holt, 1890. Pp. 605–642.

KAHNEMAN, D. Time-intensity reciprocity under various conditions of adaptation and backward masking. *J. exp. Psychol.*, 1966, **71**, 543–549.

KAHNEMAN, D. Temporal factors in vision: Some comments on terms and methods. *Psychol. Bull.*, in press.

KAHNEMAN, D., & NORMAN, J. The time-intensity relation in visual perception as a function of observer's task. *J. exp. Psychol.*, 1964, **68**, 215–220.

MURPHREE, O. D. Maximum rates of form perception and the alpha rhythm: An investigation and test of current nerve net theory. *J. exp. Psychol.*, 1954, **48**, 57–61.

PIÈRON, H. *The sensations: Their functions, processes and mechanisms.* London: Frederick Muller, 1952.

SPERLING, G. A. A model for visual memory tasks. *Hum. Factors*, 1963, **5**, 19–31.

STROUD, J. M. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1956. Pp. 174–207.

WHITE, C. T. Temporal numerosity and the psychological unit of duration. *Psychol. Monogr.*, 1963, **77**(12, Whole No. 575).

WHITE, C. T., & CHEATHAM, P. G. Temporal numerosity: IV. A comparison of the major senses. *J. exp. Psychol.*, 1959, **58**, 441–444.

(Received July 26, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 485–491

VERBAL SELF-CONTROL: THE ESTABLISHMENT OF EFFECTIVE SELF-INSTRUCTION[1]

SANDRA L. BEM

University of Michigan

3- and 4-yr.-old Ss were required to press a lever that number of times corresponding exactly to a number of lights displayed and then covered. The mean percentage of correct responses for 4-yr.-olds was 94%, while that for 3-yr.-olds was only 28% (p = .002). Once 3-yr.-olds had been trained to respond correctly in the presence of external feedback, control of their behavior was shifted to their own self-instruction by means of a fading procedure. The success of fading indicates (a) that failures by 3-yr.-olds represent a learning deficit, not a developmental deficiency, and (b) that verbal self-control can be established experimentally by means of a learning procedure analogous to that used for establishing both external stimulus control and nonverbal self-control.

The present experiment is an investigation of the ontogenetic origins of verbal self-control, the ability to generate a self-instruction, and to respond to it appropriately. Earlier research has suggested that young children are not able to regulate their own behavior by means of self-instruction. Luria (1960) reports that the typical child under 4 yr. of age is not able to respond correctly when instructed to squeeze a rubber bulb twice at the onset of a particular signal. The child "gives two, three, four, even five or six reactions, but he cannot reach a real control of pressing only twice to each signal [p. 377]."

The same child can respond correctly, however, if he is given an external signal for response termination, if, for example, he is told to terminate his behavior at the sound of a bell. He can also respond appropriately to the motor components of his own speech as signals for response termination; if, for example, he is instructed to say "Toot, toot" at the onset of each signal, each of his verbal responses activates a single corresponding motor response, and his resulting behavior pattern is correct. As soon as these *E*-induced stimuli for response termination are removed, however, the child returns to his old pattern of failure. Using the pressing of a spring-loaded lever as the dependent variable, Birch (1966) reported similar failures by American children on a number of related tasks. Luria (1961) concluded that this absence of self-control was due to a poorly developed interoceptive cue system, a system of self-generated stimuli that could provide signals for response termination in the absence of external stimuli, but which does not emerge until the child's fourth or fifth yr.

The present experiment attempts to demonstrate that such a system of self-generated cues can be established

---

[1] This research was supported in part by the Language Development Program, Center for Human Growth and Development, University of Michigan, with funds from National Institute of Mental Health Grant HD 01368-02, and in part by a summer fellowship granted to the author by the National Science Foundation. The author is grateful to David Birch for serving as advisor and to Daryl J. Bem for critical comments on the manuscript.

experimentally in 3-yr.-old children whose pretest performance indicates an absence of verbal self-control. If effective self-instruction can be established during an experimental period of 3 wk., the absence of verbal self-control would appear to result not from a developmental deficiency, but from a learning deficit. The task designed to measure verbal self-control requires the child to press a lever that number of times corresponding exactly to the number of lights presented on a display; because the child is not permitted to respond until after the lights have been covered, he must make use of self-generated stimuli, i.e., counting responses, if he is to respond correctly.

METHOD[2]

Subjects

Six 3-yr.-old and six 4-yr.-old nursery school children served as *S*s; four were male and eight were female. Two of these *S*s participated in a pilot experiment which used a slight variation of the present procedure. Three additional 3-yr.-old *S*s left the nursery school before training could be completed; only their pretest data are reported.

Apparatus

The apparatus included a panel of 12, 6-w. lights, each of which could be turned on and off independently; a rheostat which enabled *E* to dim all lights to zero simultaneously; a wooden box decorated as the face of a clown with a spring-loaded lever designed as the clown's tongue protruding from it. An 8½ × 11 in. cardboard was used to cover the lights.

Tootsie rolls, cinnamon candy, and gold stars were given to each child at the end of an experimental session. When each child completed his final session, he was given a small toy.

Procedure[3]

The *E* spent 1 day at the nursery school in order to meet the children and to become a familiar adult to them. Each *S* was then brought to a private room of the nursery school and was seated at a small table. He was shown a basket from which he was told to choose one toy to be taken home "not today, but when the game is all finished." The *S* was also told that he would be given a piece of candy and a gold star at the end of every session, that he would be able to take his toy home when he had won "a lot of stars."

After demonstrating that the lever could be manipulated, *E* instructed *S* to press it down and to let it up. The *E* then demonstrated that each lever press turned off one light. At this time, however, *S* was not given the opportunity to turn lights off.

Pretests I and II.—One, two, three, four, or five of the 12 lights were illuminated, and *S* was asked to count them, that is, to generate one number for each. The lights were then covered, and *S* was asked to repeat all of his counting responses in the absence of the lights. His performance was considered correct if he correctly repeated his prior count, whether or not that count had itself been correct. The lights were again displayed, and *S* was asked to count them once more. Finally *E* again covered them, and *S* was asked to turn them off, that is, to press the lever until he thought all the lights were off and then to stop. The *S* was given no feedback at any time during the pretests; each behavior was demonstrated before *E* asked *S* to perform it. Three behaviors were thus measured: counting in the presence of the lights, counting in the absence of the lights, and pressing in the absence of the lights. The pretest was repeated a second time. It was expected that performance would not improve from the first administration to the second, that is, taking the test would not itself be responsible for successful test performance.

Only *S*s who failed the pretest were to be eligible for training. Therefore, to insure that no amount of prompting could induce successful test performance, *E* explicitly demonstrated how successful test performance could be achieved; after counting the lights and covering them, *E* pressed and counted aloud that number of times corresponding exactly to the number of lights

---

[2] The invaluable assistance of Mary Burns of Little Farm Nursery School and Margaret Towsley of Children's Play School is gratefully acknowledged.

[3] Complete instructions to *S*s can be found in Bem (1966).

previously shown and asked *S* to do likewise. This procedure followed immediately after the second pretest.

Each pretest consisted of 15 trials, three presentations each of one, two, three, four, or five lights. The order of presentation was randomized, as well as the selection of lights to be illuminated during any single trial. Pretests were separated by at least 1 day.

The *S*s were assigned to the various training procedures described below on the basis of their pretest performance. An *S* assigned to a given training procedure completed its requirements as well as those of all subsequent training procedures, in the order specified. Each *S* remained in a given phase of training until he reached a criterion of perfect performance for 15 consecutive trials. The *S*s who reached criterion (93% correct) on all parts of the pretest were not trained further. Each phase of training was followed by a test identical to the pretest.

Each *S* attended as many sessions as necessary for his training to be completed. The number of sessions ranged from 2, for those *S*s who required no training, to 16, for those *S*s who required all training. A session typically lasted 10–15 min., ending when *S* showed signs of fatigue or boredom. It will be noted that each *S* is his own control, and that each *S* provides a complete replication of the experiment.

Verbal training.—The *S*s who did not reach a criterion of 14/15 or 93% correct on any of the three behaviors measured on the pretest were assigned to verbal training. Each of these *S*s was instructed to count the lights presented on the display, that is, to generate one counting response for each, and his response was designated by *E* as either right or wrong. If the response was wrong, *E* gave the correct response and required that *S* repeat it. Training continued until *S* reached criterion. At this point, all *S*s could count perfectly in the presence of the lights.

Verbal fading.—The *S*s who could count, but whose counting behavior in the absence of the lights did not reach criterion on the pretest, were assigned to verbal fading; this includes *S*s who had just learned to count as a result of the above procedure. Each of these *S*s was asked to count the lights presented on the display and then to repeat those same numbers. Before the second sequence of counting responses was emitted, however, *E* reduced the light intensity slightly. On each succeeding trial, the light intensity was reduced further after *S*'s initial count until the lights were no longer discriminable. Because self-generated cues, i.e., the initial sequence of counting responses, were present along with the external ones during these trials, control of the second sequence of responses should shift from the disappearing external stimuli to the available self-generated stimuli. Verbal fading should thus enable all *S*s to repeat their prior counting in the absence of the lights.

Motor training.—The *S*s whose counting behavior reached criterion both in the presence and the absence of the lights were assigned to motor training. Lights were shown to *S*, and he was asked to count them; he was further instructed to press the lever repeatedly until all the lights were off. During this phase of training, *S* could see each lever press turn off one light; since external stimuli were present throughout, this phase of training was not expected to enhance self-control.

Verbal-motor training.—Each *S* was required to count the lights and then to turn them off by counting and pressing simultaneously. Following this procedure, all *S*s have been trained to emit all the behaviors appropriate to successful test performance, but those behaviors are still controlled only by external stimuli.

Verbal-motor fading.—Each *S* was required to count the lights on the display and then to turn the lights off by counting and pressing simultaneously. Before *S* turned off the lights, however, *E* reduced the light intensity slightly. On each succeeding trial, the light intensity was reduced further after *S*'s initial count until, eventually, the intensity of the lights was so low that no discriminable environmental change accompanied *S*'s motor behavior. Because the self-generated cues were present along with the external ones during these trials, control of the motor response should shift from the disappearing external stimuli to the available self-generated stimuli, i.e., *S*'s second sequence of counting responses. Because both his verbal and his motor behavior should now be under the control of self-generated stimuli, *S* should be able to respond correctly on the test.

Pilot variations.—The pilot apparatus was identical to that of the major experiment with the exception that a panel of only four lights was used. The *S* was exposed to all phases of training in the order specified with the exception of verbal fading; during verbal-motor fading, control of both his verbal and his motor behavior was shifted simultaneously from the lights to the avail-

TABLE 1

A COMPARISON OF 3-YR.-OLDS VS. 4-YR.-OLDS: MEAN NUMBER AND PERCENTAGE OF CORRECT RESPONSES OUT OF 15 ON THE PRETEST

	3-yr.-olds ($N = 7$)[a]		4-yr.-olds ($N = 6$)		U	p (two-tailed)
	Number	Percent	Number	Percent		
Counting—Lights Present	13.9	92.4%	14.7	97.8%	10.5	
Counting—Lights Absent	11.1	74.3%	14.3	95.6%	3	.008
Pressing—Lights Absent	4.4	29.5%	9.0	60.0%	8.5	
Pressing—Lights Absent (After Prompting)	4.1	27.6%	14.2	94.4%	0	.002

[a] The *S*s from the pilot study are not included because not all four measures were taken.

able self-generated stimuli. Only three tests were administered: one before training had begun, one following verbal training, and one following the completion of all training. Each test consisted of 16 trials. Two behaviors were measured, i.e., counting in the presence of the lights and pressing in their absence.

Results

Pretest Data—Developmental Trends

Counting—lights present.—Three-yr.-old children are able to count. Five out of seven *S*s responded correctly on more than 93% of the trials. The number of correct responses by 4-yr.-olds was not significantly greater according to a two-tailed Mann-Whitney U test ($U = 10.5$, $p \leqslant .18$), although five out of six *S*s responded perfectly on all trials.

Counting—lights absent.—In the absence of external stimuli, 4-yr.-olds remained able to generate the appropriate number of counting responses. As Table 1 indicates, they correctly repeated their prior verbal responses after the lights had been covered on 95.6% of the pretest trials. Their performance was significantly better than that of the 3-yr.-olds ($U = 3$, $p = .008$) who were able, nevertheless, to respond correctly on 74.3% of the trials.

Pressing—lights absent.—Prior to any training, neither group was consistently able to terminate its motor behavior appropriately in the absence of external cues for response termination, that is, when the lights were covered. Three-yr.-olds pressed the lever the correct number of times on 29.5% of the trials, while 4-yr.-olds did so on 60% of them. The difference between the groups was not significant ($U = 8.5$, $p = .10$).

Thus, most *S*s remained unable to control their motor behavior appropriately, despite the fact that they were able to repeat their previous counting responses in the absence of external stimuli. This indicates that *S*s do, in fact, have available the responses that could serve as mediators of their subsequent motor behavior, but that these responses do not yet function as controlling variables.

Pressing and counting—lights absent.—Explicit demonstration by *E* enabled 4-yr.-olds to perform appropriately on the test. After *E* prompted them by pressing and counting aloud the correct number of times, these *S*s responded correctly on 94.4% of the trials. Because correct performance could be induced, none of the six 4-yr.-olds participated in training. The new instruction was not able, however, to effect improved performance in the 3-yr.-olds who responded correctly on only 27.6% of the trials. The pretest data considered alone, then, are consistent with the hypothesis that verbal self-control is a developmental phenomenon. It remains to be demonstrated that correct performance by the 3-yr.-olds depends crucially not upon a particular developmental level, but upon a learning procedure.

Results of Training

Effects of entire procedure.—At the completion of training, all *S*s were able to respond correctly. The solid lines in Fig. 1 indicate that five out of six *S*s responded perfectly, pressing the lever the correct number of times on every trial.[4] Only one error was made out of a total of 92 trials.

These data clearly indicate that the training procedure is sufficient for successful test performance, for self-control of response termination. They further demonstrate that the reported absence of mediation is not due to a developmental deficiency, but to a learning deficit. If an *S* is trained to respond to the self-generated stimuli of his own responses, that is, if his responses are forced to function as discriminative stimuli, that *S* is able to control his own behavior.

Effects of pretesting.—Figure 1 indicates that taking the pretest is not itself responsible for correct responding; none of the four *S*s who took both pretests improved from the first administration to the second (*S*s 3, 4, 5, & 6).

Effects of verbal training.—Training *S*s only to count the lights pre-

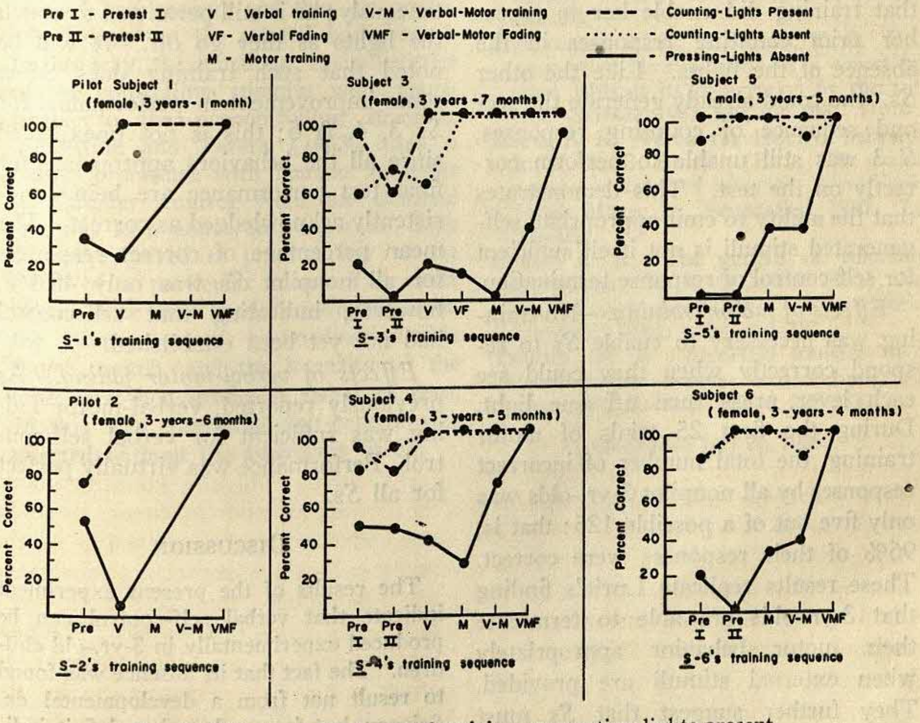


FIG. 1. Effects of all phases of training on counting-lights present, counting-lights absent, and pressing-lights absent.

---

[4] Figure 1 does not include the data of 3-yr.-olds whose pretest data have been reported, but who left the nursery school before training could be completed.

sented on the display did not effect improved test performance; the performance of three out of four *S*s declined (*S*s 1, 2, & 4), while that of the fourth improved, but only to a previously achieved level (*S* 3). Out of 15 trials, the mean-percentage correct was only 22.6% following verbal training. Clearly, verbal training does not itself produce successful performance on the test. As the dashed lines in Fig. 1 indicate, however, it did enable *S*s 1, 2, 3, & 4 to count perfectly Because *S*s 5 and 6 reached criterion on the pretest, they did not participate in verbal training.

Effects of verbal fading.—Because three out of four nonpilot *S*s were able to repeat their prior counting responses on the pretest, whether or not those responses had themselves been correct, only *S* 3 underwent verbal fading; as indicated by the dotted line in Fig. 1, that training did enable her to repeat her prior counting responses in the absence of the lights. Like the other *S*s, who could already generate the second sequence of counting responses, *S* 3 was still unable to perform correctly on the test. This demonstrates that the ability to emit appropriate self-generated stimuli is not itself sufficient for self-control of response termination.

Effects of motor training.—No training was necessary to enable *S*s to respond correctly when they could see each lever press turn off one light. During the first 25 trials of motor training, the total number of incorrect responses by all nonpilot 3-yr.-olds was only five out of a possible 125; that is, 96% of their responses were correct. These results replicate Luria's finding that 3-yr.-olds are able to terminate their motor behavior appropriately when external stimuli are provided. They further suggest that *S*s must learn to replace those stimuli with self-generated ones if they are to respond correctly on the test.

It will be noted that motor training improved the test performance of *S* 5 and *S* 6. Prior to motor training, both of these *S*s had been emitting a number of responses for each trial far in excess of the number of lights presented. Motor training did not teach them to respond with the appropriate number of responses, but only to approximate the appropriate range. Accordingly, the percentage of correct responses for these *S*s increased, but only to 36.7%. Seventy-five percent of all *S*s' responses remained incorrect; these data clearly indicate that the entire procedure up to and including motor training is not sufficient for self-control of response termination.

Effects of verbal-motor training.—In verbal-motor training, *S* turns lights off by pressing and counting simultaneously; he is still permitted to watch the lights as they go off. It will be noted that such training does effect some improvement in performance for *S*s 3, 4, & 6; this is not unexpected since all the behaviors appropriate for final test performance are being consistently acknowledged as correct. The mean percentage of correct responses for all nonpilot *S*s was only 48.3%, however, indicating that self-control had not yet been established.

Effects of verbal-motor fading.—As previously reported, verbal-motor fading was sufficient for verbal self-control. Performance was virtually perfect for all *S*s.

DISCUSSION

The results of the present experiment indicate that verbal self-control can be produced experimentally in 3-yr.-old children. The fact that its absence was found to result not from a developmental deficiency, but from a learning deficit indicates that speech can gain new functional significance by means of a learning procedure and, further, that a learning procedure may be necessary for the natural emergence of verbal self-control. While this does not imply that a child of any age can be trained to obey his self-instruction, it does emphasize the importance of learning in the establishment of effective self-instruction, and it therefore has direct bearing on the traditional question of the relative contributions of development and learning toward the natural emergence of verbal self-control.

The crucial component of the learning procedure used in the present experiment has become known as a fading procedure and was clearly demonstrated in an experiment by Terrace (1963). In his experiment, control of a pigeon's discrimination was shifted with no errors from the background color of a pecking disc to the orientation of a line on that disc. This was accomplished by superimposing the noncontrolling stimulus, the line, onto the controlling stimulus, the colored background, and then by gradually fading out the colored background; in this way, the functional characteristics of the controlling stimulus were transferred to the noncontrolling stimulus. Hefferline and Perera (1963) used a fading procedure with humans in order to transfer control from an external stimulus to a nonverbal response. Each *S* was required to press a key at the onset of a particular tone. The onset of the tone was made contingent upon the occurrence of a nondiscriminable thumb twitch, and the intensity of the tone was then gradually faded to zero. In this way, nonverbal self-control was effected without the awareness of *S*s.

The fact that a fading procedure was able to shift control in the present experiment from an external stimulus to a verbal self-instruction indicates not only that speech can gain new functional significance by means of a learning procedure, but also that functionally similar procedures are effective in establishing external-stimulus control, nonverbal self-control, and verbal self-control. Further, the fact that fading appears to be effective for establishing self-control in the laboratory suggests that some analogue to it may be the real world mechanism whereby speech becomes "interiorized" (Luria, 1960, p. 382) and verbal self-control emerges.

REFERENCES

BEM, S. L. Verbal self-control: The establishment of effective self-instruction. *Technical Report No.* 12, 1966 University of Michigan, Grant HD 01368-02, Center Human Growth Dev., National Institute of Mental Health.

BIRCH, D. Verbal control of nonverbal behavior. *J. exp. Child Psychol.,* 1966, **4**, 266–275.

HEFFERLINE, R. F., & PERERA, T. B. Proprioceptive discrimination of a covert operant without its observation by the subject. *Science,* 1963, **139**, 834–835.

LURIA, A. R. Verbal regulation of behavior. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy, Jr., Foundation, 1960. Pp. 359–423.

LURIA, A. R. The genesis of voluntary movements. In N. O'Connor (Ed.), *Recent Soviet psychology.* New York: Liveright, 1961. Pp. 165–185.

TERRACE, H. S. Errorless transfer of a discrimination across two continua. *J. exp. Anal. Behav.,* 1963, **6**, 223–232.

(Received July 26, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 492–499

ASYMMETRIC-REINFORCING EVENTS IN PROBABILITY LEARNING[1]

HARRY L. MADISON AND PATRICK A. BOUDEWYNS
University of Wisconsin, Milwaukee

Predicting occurrence of 1 of 2 reinforcing lights (symmetric situation) was compared with predicting occurrence and nonoccurrence of 1 light (asymmetric situation) in 2 experiments. The situations differed significantly, with Ss showing preference for predicting occurrence over nonoccurrence. 3 statistical-learning hypotheses were evaluated. A modified Estes and Straughan (1954) model, assuming occurrence to be more reinforcing (more elements conditioned per trial) than nonoccurrence, best described the differences.

Early two-choice probability learning studies (Grant, Hake, & Hornseth, 1951; Humphreys, 1939) required *S*s to predict whether a reinforcing light would or would not occur on each trial, here called an asymmetric situation. Estes and Straughan (1954) required *S*s to predict which one of two reinforcing lights would occur, a symmetric situation. It has been assumed (Estes, 1964, p. 104) that nonoccurrence of the light has the same reinforcing effect upon predicting nonoccurrence as does occurrence of the light upon predicting occurrence, and therefore the symmetric and asymmetric situations are equivalent.

The present study arose from an experimental comparison of the two situations. A preliminary experiment was conducted, using three groups of 30 *S*s over 100 probability learning trials, with π, the probability of an E_1 reinforcing event, equal to .7. For Group I, the symmetric group, the A_1 response was predicting one of the two lights, which occurred with probability π, and A_2 was predicting the other light, which occurred with probability $1 - \pi$. For Group II, an asymmetric group, A_1 was predicting the occurrence of the one light, which occurred with probability π, and A_2 was predicting its nonoccurrence, this event having probability $1 - \pi$. Group III, another asymmetric group, was treated the same as Group II, except that A_1 was predicting nonoccurrence and A_2 was predicting occurrence of the one light, events having probabilities π and $1 - \pi$, respectively. The right-left positions of E_1 events were counterbalanced within each group. There was a significant difference in total A_1 responses, $F\ (2, 84) = 5.02$, $p < .01$, and in the number of A_1's in the last block of 20 trials, $F\ (2, 84) = 3.32$, $p < .05$. The groups were ordered III < I < II in both measures. These results suggested that occurrence and nonoccurrence of a light do not have the same reinforcing effects, and therefore the symmetric and asymmetric situations are not equivalent.

However, Grant et al. (1951) reported no difference in results between *S*s instructed to make an overt response when they predicted the one light would

---

[1] This study was supported by grants from the University of Wisconsin Graduate School. Part of Exp. I was read at the Midwestern Psychological Association meeting, Chicago, 1962, with J. Kvietins. Experiment II is based upon the MS thesis of the second author, and part of it was read at the American Psychological Association meeting, New York, 1966. The assistance of T. D. Beck and T. P. Theander in a preliminary experiment is acknowledged.

occur, compared with *S*s responding overtly when they predicted it would not occur. Gerjuoy, Gerjuoy, and Mathias (1964), using apparatus and procedures comparable to Groups II and III of our preliminary experiment, found no difference between the two. Neither study included a symmetric situation.

In view of these discrepant results, three hypotheses about differences between the symmetric and asymmetric situations are presented in this paper, formulated in terms of statistical association theory (Estes & Burke, 1953), along with experiments designed to test them.

Hypothesis I (*H-I*).—The symmetric and asymmetric situations are equivalent, and the Estes and Straughan (1954) statistical learning model accounts for the main features of the data. This model leads to the well-known expression

$$\bar{P}_n = \pi - (\pi - P_o)(1 - \theta)^n \quad [1]$$

where $\overline{P}_n$ is the expected probability of an A_1 response following Trial n, π is the probability of an E_1 reinforcing event, P_o is the initial probability of an A_1, and θ is the proportion of stimulus elements sampled on each trial.

Hypothesis II (*H-II*).—Occurrence of the one light in the asymmetric situation is more reinforcing than its nonoccurrence. This is interpreted as causing a proportion V of sampled stimulus elements to become conditioned to predicting the light on a trial where the light does *not* occur, while all sampled elements become conditioned to predicting the light on a trial where it *does* occur. This assumption may be expressed as

$$P_{n+1} = (1 - \theta)P_n + \theta, \quad [2]$$

where P_{n+1} is the probability of predicting the light following a trial where it does occur, and

$$P_{n+1} = (1 - \theta)P_n + V\theta, \quad [3]$$

following a trial where the light does not occur. These equations have the solution

$$\bar{P}_n = \pi + V(1 - \pi) - [\pi + V(1 - \pi) - P_o](1 - \theta)^n, \quad [4]$$

where π is the probability of occurrence of the light. If no-light trials occur with probability π, Equations 2 and 3 have the solution

$$\bar{P}_n = \pi - V\pi - [\pi - V\pi - P_o](1 - \theta)^n, \quad [5]$$

where $\overline{P}_n$ is the probability of predicting no light following Trial n. Equations 4 and 5 reduce to Equation 1 for the symmetric situation, when $V = 0$.

Hypothesis III (*H-III*).—Differences between the symmetric and asymmetric situations are initially as described by H-II, but V decreases over trials. The rationale for H-III arises from consideration of the temporal relationships among the experimental events in the two situations. The usual procedure in the symmetric situation is to have one of the two reinforcing lights occur at a fixed time following the signal which begins the trial. However, in the asymmetric situation it is only when the one light occurs that there is a definite reinforcing stimulus at this time. Complete information about the outcome of a no-light trial is given by the occurrence of the signal for the next trial, which occurs at a later time. We might thus expect S to be relatively uncertain about the outcome of a no-light trial, at least until he has formed an adequate temporal discrimination, at which time nonoccurrence of the one reinforcing light is as definite an event as is its occurrence. The rate at which the symmetric and asymmetric situations tend to become indistinguishable will depend upon the particular relationship between V and n which is assumed; since H-III admits of so many mathematical formulations, none will be given unless the data compel its retention.

In order to evaluate these hypotheses, two experiments are presented for which they yield differential predictions. Our first experiment compares symmetric and asymmetric situations during probability learning and reversal. Consider three groups treated as in our preliminary experiment during a training series, with the probability of an E_1 reinforcement equal to π, followed by a reversal series with this probability changed to $1 - \pi$. According to H-1, the groups will show identical learning curves, described by Equation 1, with asymptotic A_1 response probabilities equal to π for the training series and to $1 - \pi$ for the reversal series.

According to H-II, Group I, the symmetric group, is described by Equation 1, since $V = 0$. Group II, the asymmetric group whose A_1 response is predicting the light, is described by Equation 4, with training and reversal asymptotes equal to $\pi + V(1 - \pi)$ and $1 - \pi + V\pi$, respectively. Group III, the asymmetric group whose A_1 response is predicting no light, is described by Equation 5, with training and reversal asymptotes equal to $\pi - V\pi$ and $1 - \pi - V(1 - \pi)$, respectively. Therefore, the asymptotic A_1 response probabilities of the three groups for the training series will be ordered III < I < II, since

$$\pi - V\pi < \pi < \pi + V(1 - \pi), \quad [6]$$

for $V > 0$, $0 < \pi < 1$. It can be seen that the same order will occur for the reversal series. By comparing Equations 1, 4, and 5, in the manner of Equation 6, it can be shown that the same order will hold for A_1 response probabilities following all trials in a series, under two conditions of interest: first, when the P_o's for the three groups are equal, a condition which can be approximated at the beginning of training by counterbalancing the side of the E_1 reinforcement; and second, when the P_o's are ordered III < I < II, a condition which will obtain at the beginning of the reversal series. Since the total A_1 responses in a series is given by the sum of the A_1 response probabilities over trials, the same order will hold for total A_1's. Therefore, H-II predicts different learning curves for the three groups, with asymptotes and total A_1's ordered III < I < II in both training and reversal.

According to H-III, by previous arguments, A_1 response probabilities will be ordered III < I < II during early trials, but this ordering will tend to disappear over trials as V approaches zero throughout training and reversal. Therefore, given sufficient trials, H-III predicts that differences between the symmetric and asymmetric situations will disappear.

EXPERIMENT I

Method

Subjects and apparatus.—The *S*s were 198 students from beginning psychology courses, assigned at random to three groups of 66 *S*s each. All *S*s were tested individually. The *E* and *S* sat on opposite sides of a table, separated by a large opaque screen. On *S*'s side two telegraph keys were mounted 17 in. apart at the base of the screen; upon the screen were mounted three neon panel lights. One of these lights, which served as the signal for *S* to guess, was centered 17 in. above the telegraph keys, at about *S*'s eye level. Each of the two remaining lights, the reinforcing lights, was mounted directly above one of the telegraph keys. For Group I both reinforcing lights were exposed; for Groups II and III one of them was covered.

Procedure.—Group I *S*s were read instructions requesting that they respond to the signal on each trial by operating the telegraph key corresponding to the reinforcing light they expected to follow. For each *S*, one of the two reinforcing lights was randomly designated as E_1, and the other as E_2. For half the *S*s, E_1 was the right-hand light, and was the left-hand light for the others.

Group II *S*s were read similar instructions, except that they were requested to guess whether the one reinforcing light would occur following the signal, in which case the key beneath the light was to be operated, or whether the light would not occur, in which case the other telegraph key was to be operated. For each *S*, occurrence of the one reinforcing light was E_1. It was located on the right for half the *S*s and on the left for the others. Nonoccurrence of the light was E_2.

Group III *S*s were treated the same as Group II, except that nonoccurrence of the one reinforcing light was E_1. Its location was also counterbalanced.

Four practice trials were given, with the reinforcing events presented in the order E_1, E_2, E_2, E_1 for half the *S*s in each subgroup, and E_2, E_1, E_1, E_2 for the others. Following the practice trials, 120 training trials were given, with $\pi = .7$. The training trials were followed without interruption by 120 reversal trials, with π changed to .3.

On each trial the signal light was lighted for 1.0 sec., followed 2 sec. later by a reinforcing light (if one was to occur) which remained on for 1 sec. The time interval between signal lights was 5 sec. These intervals were timed by *E* with a stopwatch.

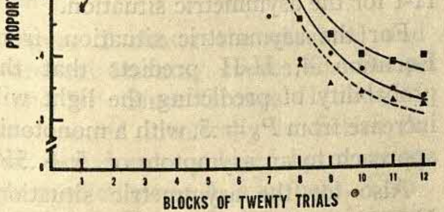


FIG. 1. Observed and theoretical curves representing mean proportion of A_1 responses per 20-trial blocks for training and reversal.

Results

The mean proportions of A_1 responses over blocks of 20 trials are shown in Fig. 1. For the training series, both the total and terminal (last 40 trials) A_1's were ordered III < I < II, with significant differences, F (2, 192) = 20.51, $p < .001$, and 12.60, $p < .001$, respectively. However, although the ordering of groups was as predicted by H-II, a Newman-Keuls test of the terminal levels failed to show a significant difference between Groups I and II, suggesting that H-III may be true, at least where the one light occurs most often. This was substantiated by a significant Groups × Trial Blocks interaction, F (5, 640) = 3.01, $p < .025$, when Groups I and II were compared. Groups II and III did differ significantly in both total and terminal A_1's, contrary to the findings of Gerjuoy et al. (1964) in a comparable asymmetric situation.

The reversal results are also shown in Fig. 1. Both total and terminal A_1's were ordered III < I < II, as predicted by H-II, with significant differences, F (2, 192) = 5.91, $p < .005$, and 3.43, $p < .05$, respectively. However, a Newman-Keuls test failed to show a significant difference between the terminal levels of Groups I and III. Since Group III during reversal was treated the same as Group II during training, this also suggested that H-III may be true when the one light occurs most often. There was also a Groups × Trial Blocks interaction, F (5, 640) = 2.21, p slightly greater than .05, when Groups I and III were compared. Groups II and III did differ significantly in total and terminal A_1's, indicating that the difference shown during training between predicting the light and predicting no light persisted during reversal.

The theoretical training curve for Group I was obtained by estimating $\theta = .0242$ from the total A_1's and the proportion of A_1's in the first trial block, a procedure described by Estes and Straughan (1954). The theoretical training curve for Group II was obtained under H-II,[9] using the total A_1's and observed A_1's in the first block, along with θ from Group I, to estimate $V = .0686$. Comparison of the observed and theoretical total and terminal A_1's for Group II showed little difference, t (65) < 1.0, and 1.50, $p > .20$. The theoretical training curve for Group III was obtained by using the observed A_1's in the first block, along with the previously estimated θ and V. As can be seen, it is not a good fit. The predicted total A_1's were greater than observed, t (65) $= 1.90$, $.05 < p < .10$, and the predicted terminal level was higher than observed, t (65) $= 3.18$, $p < .01$. Of course, V may be estimated from Group III and used to predict Group II, with a resulting poor fit for Group II.

The theoretical reversal curves shown in Fig. 1 were obtained under H-II without reference to the reversal data, using the θ and V estimated from the training data. Early reversal learning occurred more rapidly than predicted for all groups. Comparisons of observed and predicted A_1's in the first 40 reversal trials for Groups I, II, and III yielded t (65) $= 3.51$, $p < .001$; 2.69, $p < .01$; and 2.12, $p < .05$, respectively. Agreement between prediction and observation was better during the later reversal trials, and comparisons of observed vs. theoretical terminal A_1's for each of the groups yielded t (65) < 1.32, $p > .10$.

The data were also analyzed for the effects of left vs. right locations of E_1. Although there appeared to be an overall right-response preference in both training and reversal, the differences in total A_1's between E_1 right and E_1 left were not significant. Gerjuoy et al. (1964) reported a significant right-response preference in an asymmetric situation comparable to our Groups II and III. Within groups, the only significant difference was a right-response preference during reversal for Group I, t (64) $= 2.28$, $p < .05$. This finding is contrary to Miller (1966), who used only right-handed Ss, and found that a right-response preference in training disappeared during reversal in a symmetric situation comparable to Group I. However, in assigning Ss in Exp. I we did not control for handedness, and any differential effects due to the (presumably few) left-handed Ss were confounded with the other variables. This has apparently been true for many other human probability-learning studies.

The results of Exp. I permit the rejection of H-I, but they do not provide a clear choice between H-II and H-III. A second experiment was conducted, using only right-handed Ss, comparing symmetric and asymmetric groups over a series of trials with $\pi = .5$. Assuming $P_o = .5$, our three hypotheses predict a straight, horizontal line, $\overline{P}_n = .5$, for the symmetric situation, i.e., no change in A_1 response probabilities are predicted, and thus no learning will occur. The same result is predicted by H-I for the asymmetric situation.

For the asymmetric situation, from Equation 4, H-II predicts that the probability of predicting the light will increase from $P_o = .5$, with a monotonic approach to an asymptote of $.5 + .5V$.

Also for the asymmetric situation, H-III predicts that the probability of predicting the light will at first increase from $P_o = .5$, and then decrease toward .5 as V approaches zero, a nonmonotonic function.

As a further evaluation of our hypotheses, an additional series of trials was given with $\pi = .7$, comparable to the training series in Exp. I. Under this condition, H-I predicts no difference between the symmetric and asymmetric situations; H-II predicts that the asymptotic and total A_1's will be ordered III < I < II, as will be the P_0's because of learning in the previous series; and, H-III predicts that differences between the symmetric and asymmetric situations will continue to decrease. This additional series was also given to obtain more information about right-response preference in probability learning, especially to examine Miller's (1966) suggestion that a preexperimental right-preference disappears over training.

EXPERIMENT II

Method

Subjects and apparatus.—The *S*s were 180 right-handed, beginning psychology students, randomly assigned to three groups of 60 *S*s each. Each *S* was tested individually. Handedness was determined after *S* was tested; the data of left-handed *S*s was not analyzed. The apparatus was the same as that used in Exp. I, with the addition of electronic timing of the intervals, and tape-recorded instructions.

Procedure.—Except for the π values, the procedure was the same as in Exp. I. A series of 120 trials (Series A) was given with $\pi = .5$, followed without interruption by 120 more trials (Series B) with $\pi = .7$.

Results

Figure 2 shows the observed and theoretical learning curves for Series A and B. Since Groups II and III were treated identically in Series A, their data were combined by counting light predictions as A_1's for both. Comparison of Group I (symmetric) vs. Groups II and III (asymmetric) in Series A showed a difference in both total and terminal A_1's, t (178) = 7.09, $p < .001$, and 5.81, $p < .001$, respectively. The symmetric and asymmetric situations differed at $\pi = .5$, as predicted by H-II, with *S*s showing a preference for predicting the occurrence rather than the nonoccurrence of the light in the asymmetric situation.

Comparisons of Groups I, II, and III in Series B showed significant differences in total and terminal A_1's, F (2, 174) = 31.56, $p < .001$, and 26.93, $p < .001$, respectively. Newman-Keuls tests showed significant differences between the groups on both measures, which were ordered III < I < II, as predicted by H-II. The Groups × Trial Blocks interactions were not significant for either total A_1's or for the number of A_1's in the last four blocks of 10 trials, Fs < 1.0, so there was no evidence that this ordering was breaking down as predicted by H-III

The theoretical curves were obtained by first estimating $\theta = .0145$ from the Group I, Series B data, and using it to obtain the Group I, Series B curve as in Exp. I. The Group I, Series A curve is simply a straight line with $\overline{P}_n = .5$. The Groups II and III, Series A data were then used to estimate $V = .142$ from the mean observed proportion of A_1's over the last

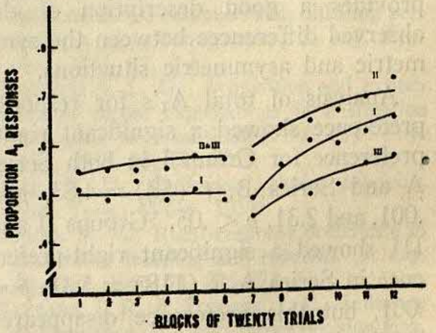


FIG. 2. Observed and theoretical curves representing mean proportion of A_1 responses per 20-trial blocks for each series.

40 trials, assuming the groups were at an asymptote of $\pi + V(1 - \pi)$, from Equation 4. Since little change in proportion of A_1's occurred after the early trials, this assumption seemed reasonable. Both θ and V were then used to obtain the Series A curve for Groups II and III, under H-II. The mean proportion of A_1's over the last 40 trials of Series A for Groups II and III was used as the theoretical initial probability of a light prediction at the beginning of Series B for these groups, and together with θ and V was used to obtain their Series B curves. Thus, none of the Group II or III Series B data were used in obtaining the theoretical curves.

Comparisons between the observed and theoretical total and terminal A_1's for the five curves in Fig. 2 yielded t (59) $< .85$, $p > .40$, in each case. Agreement between observed and theoretical learning curves seemed to be good. While such a conclusion based upon failure to reject the null hypothesis may properly be questioned, with 60 *S*s per group the *t* tests should be reasonably sensitive to differences between theory and observation. The groups were correctly ordered according to H-II, and the observed differences among the groups were significant. It was concluded that H-II provides a good description of the observed differences between the symmetric and asymmetric situations.

Analysis of total A_1's for response preference showed a significant right-preference for Group I in both Series A and Series B, t (58) $= 4.83$, $p < .001$, and 2.31, $p < .05$. Groups II and III showed a significant right-preference in Series A, t (118) $= 5.31$, $p < .001$, but this preference disappeared by the last block of 20 trials, $t < 1.0$, and no significant right-preference was shown by Groups II or III in Series B, *t*'s < 1.0. While the results for the asymmetric groups were consistent with Miller's (1966) suggestion that extinction or adaptation of a preexperimental response preference occurs, the results for the symmetric group were contrary to it. It was as if preference for predicting the one reinforcing light in the asymmetric situation overcame a right-response tendency, while the tendency persisted in the symmetric situation. Further investigation will be needed to examine this question.

DISCUSSION

The two experiments presented support the idea that the symmetric and asymmetric situations in two-choice human probability learning, previously considered equivalent, yield different results. A reasonable account of this difference is obtained by assuming that occurrence of a single reinforcing light is more reinforcing, i.e., more sampled elements are conditioned per trial. The differences predicted under H-II depend upon the proportion of sampled stimulus elements assumed to become conditioned to predicting the light on trials when it does not occur. These predicted differences can be rather small, and therefore easily obscured by sampling error. For example, if we take $V = .0686$ and $\pi = .7$ from Exp. I, the predicted asymptote for an asymmetric group for whom occurrence of the light is E_1 will be $\pi + V(1 - \pi) = .72058$, which is only 2.94% greater than .7, the asymptote for a symmetric group. If nonoccurrence is E_1, the predicted asymptote will be $\pi(1 - V) = .65198$, which is 6.86% less than .7, and less likely to be obscured by sampling error. This may be what occurred in both training and reversal in Exp. I, although the significant Groups × Trial Blocks interactions which were found argue that it was not that simple. The result may have been due to some left-handed *S*s.

In both experiments the predicted asymptotes were not reached, except perhaps for the $\pi = .5$ condition in Exp. II, and the significant differences observed between the symmetric and asymmetric situations may still break down over more extended training. Our results, particularly in Exp. II, suggest that if this is happening it is a very slow process.

The results of both experiments would not seem to support the idea, expressed in H-III, that omission of an experimental reinforcing event on no-light trials has a differential effect because of any temporary uncertainty about the outcome of the trial. For example, by the time the last 40 trials occurred, *S*s in Group III of Exp. I had experienced this 108 times, and *S*s in Exp. II, 116 times. Failure of the reinforcing light to occur should by that time have been about as certain an event as its occurrence, but the terminal performances of the two asymmetric groups in both experiments were very different.

While the parameter, *V*, has been given meaning within the context of statistical learning theory, a consideration of its psychological meaning is in order. By coincidence, the same symbol was used by Luce (1959) to stand for the vigor of the unconditioned response. One could interpret H-II in this way, assuming the proportion of sampled elements conditioned to a response to be related to the vigor of the unconditioned response which reinforces it, in the present situation presumably an implicit verbal response (Estes & Straughan, 1954). However, there is no obvious way to measure this quantity directly. Perhaps comparisons of GSRs or EMGs during light and no-light trials would yield results of interest. One could also hypothesize that the vigor of the conditioned response (guessing behavior and key press) effects the proportion of sampled elements conditioned to the response. However, in the present study it was observed that, if anything, more vigor was shown by *S*s in the asymmetric situation when they were predicting nonoccurrence of the light, with *S*s pounding harder on the key, shuffling their feet more, etc. These were only casual observations, but if such behavior were to be verified through systematic measurements this would argue against the implications of the second hypothesis above. On the other hand, such behavior could be considered indicative of an emotional effect of failure of the reinforcing light to occur, with the implication that this interferes with the learning of the response of predicting no light. Finally, one direction further research could take is indicated in a probability-learning study by Nazzaro and Todorov (1966), who used two reinforcing lights of unequal luminance. They found, at $\pi = .75$, that *S*s tended to make more predictions of the bright than the dim light, with predictions of lights of equal brightness falling in between. Since the present study could be thought of as representing an extreme case of differences in brightness, i.e., no light at all vs. an easily perceived one, further theoretical and experimental investigation along this line may be promising.

REFERENCES

Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Gerjuoy, I. R., Gerjuoy, H., & Mathias, R. Probability learning: Left-right variables and response latency. *J. exp. Psychol.*, 1964, **68**, 344–350.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Miller, M. E. Right-response preference in probability learning and reversal. *J. exp. Psychol.*, 1966, **71**, 776–778.

Nazzaro, J. R., & Todorov, J. C. Influence of luminance on a two-choice decision task. *J. exp. Psychol.*, 1966, **71**, 696–699.

(Received July 29, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 500–505

RELATION BETWEEN STIMULUS RECOGNITION AND PAIRED-ASSOCIATE LEARNING[1]

EDWIN MARTIN

University of Michigan

The hypotheses tested were that association activation, as evidenced by correct responding, is contingent upon stimulus recognition but that stimulus recognition does not depend upon the existence of associated responses. The data support both hypotheses. Using the study-test method, Ss were given practice on 12 trigram-number pairs where 6 of the pairs remained intact and 6 were re-paired anew on each study trial. On each test trial, the 12 study-trials trigrams were randomly intermixed with 12 new filler trigrams. The S had first to make a stimulus-recognition response and then to emit a number response.

The principal question to which the present research is directed is: What role does stimulus recognition play in paired-associate learning? A related question is: Does association of a response with a stimulus affect the recognizability of that stimulus?

It was argued in a previous paper (Martin, in press) that "regardless of the current status of an S-R association, the activation of that association, and hence the occurrence of the response event R, has as a necessary antecedent recognition of the stimulus event S." It was shown that irrespective of the number of times a response had been given correctly on previous trials, the probability of its again being given correctly falls to the level of chance whenever the learner fails to recognize the stimulus.

The strong inference of the necessity of recognition for association activation following from the work just mentioned is somewhat marred, however, by the fact that the reverse contingency, that recognition depends upon the existence of an association, could not clearly be ruled out. The experiment reported here shows this reverse contingency to be highly unlikely and further affirms the necessity inference.

METHOD

Design and materials.—In this experiment, *S*s were given 12 trials on a single, aurally presented list of 12 trigram-number pairs according to the study-test method. The stimuli were CCCs of 45–55% association value according to Witmer (1935); the responses were the numbers, 1, 2, and 3. Of these 12 pairs, there were six each of what will be called Types C and R. Type C: The six Type-C pairs were consistently paired throughout learning. The numbers, 1, 2, and 3, were each paired with the same two trigrams over the 12 study trials. Type R: The six Type-R pairs were re-paired anew on each study trial. From trial to trial, the numbers 1, 2, and 3 were variably paired with the trigram stimuli according to a schedule that was random except for the restrictions that (*a*) over the 12 study trials each number was paired with each trigram equally often and (*b*) within each study trial each of the numbers was used exactly twice.

Actually, six such 12-pair lists were constructed so that every trigram entered into a Type-C pair on three of them and into a Type-R pair on the other three. Further, over the three lists in which a trigram entered into a Type-C pair, it was paired once with each of the three number responses.

On test trials, 24 trigrams were presented.

---

[1] This research was supported by the Advanced Research Projects Agency, Department of Defense, monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1235 with the Human Performance Center, Department of Psychology. The data were collected by Louis Barr Hewitt.

To each of these *S* had to respond first with either a recognition or a nonrecognition response and then with one of the numbers, 1, 2, and 3. Of the 24 trigrams presented, 12 were the stimuli presented on study trials; while the remaining 12, which were also CCCs of 45–55% association value according to Witmer (1935), were new on each test trial. The 12 study-trials stimuli and the 12 new filler trigrams were intermixed in a different random order on every test trial.

All materials were prerecorded on tape and presented to *S*s over earphones. On the desk in front of *S* was a two-button panel with one button marked "new" and the other marked "old," thereby providing the mechanics for testing *S*'s recognition of trigrams on test trials.

Procedure.—Each *S* was told that study and test trials would alternate regularly, beginning with a test trial; that on study trials he was to listen silently as the 12 pairs were presented through the earphones at a 2-sec. rate; and that on test trials the 12 study-trials stimuli plus 12 new trigrams would be intermixedly presented at a 6-sec. rate. The test-trials task was described to him as follows:

> As you hear each trigram you must do two things. First, simply decide whether or not it is one that occurred during the study trial. If you think it did, press the button marked "old"; if it seems like a new one, press the button marked "new." Second, you must say aloud a number. When the trigram presented is one that occurred during the study trial, try to say the number that went with it. If you can't remember what it was, guess. When the trigram is a new one, simply choose any of the numbers, 1, 2, or 3, and say it aloud. Thus for each of the trigrams read to you on the test trial, press one of the buttons and say aloud a number. You must do both for each trigram.

Further, *S* was advised (in so many words) of the two kinds of study-trials pairs, Types C and R.

There was a preliminary test trial, immediately preceding the first study trial, on which *S* responded "new" and guessed a number to every trigram. Altogether, then, there were 12 study and 13 test trials; the sequence of alternating study and test trials began and ended with a test trial. The intertrial interval was 5 sec., during which time *S* was alerted as to whether the upcoming trial was to be a study or a test trial.

Subjects.—The *S*s were 48 junior and senior female University of Michigan students who volunteered for paid ($1.25) participation. Eight *S*s were randomly assigned to each of the six ostensibly equivalent lists.

RESULTS

Recognition.—A recognition is defined as pressing the "old" button on a test trial when the trigram presented is a study-trials stimulus. Proportion of recognitions over trials, with type of pair as the parameter, are shown in the upper part of Fig. 1. Collapsing over trials, proportion of recognitions of Types C and R stimuli are .829 and .826, respectively. These latter two proportions are each based on (48 *S*s) × (6 stimuli per type) × (12 trials) = 3,456 experimental events.

There are two properties of these data worth noting. (*a*) Stimulus recognition is not contingent upon pair

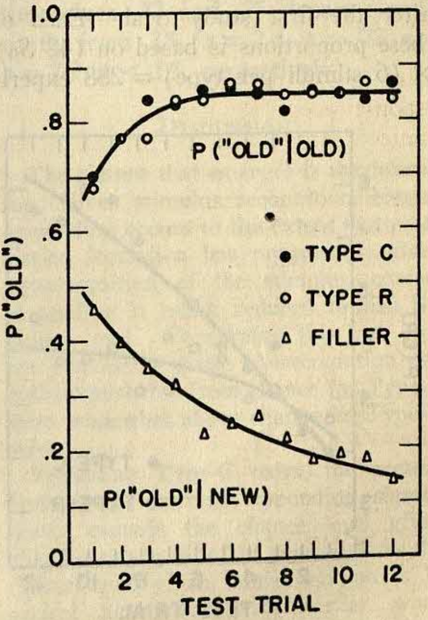


FIG. 1. Proportion "old" responses to trigrams that were old (study-trials stimuli, Types C and R) and new (fillers). (The continuous curves are visual approximations.)

type. In other words, recognition of Type-R stimuli is not impaired by the fact that trial-to-trial S-R consistency is absent; or, conversely, the recognition of Type-C stimuli is not facilitated by the fact that stable associations can be attached to them. (*b*) Beyond Trial 5, there is no further improvement in recognition; there appears to be an asymptote at .85.

Proportion of false-positive responses over trials are shown in the lower part of Fig. 1. A false-positive response is defined as pressing the "old" button on a test trial when the trigram presented is a filler trigram. Note that false-positive responding continues to drop off even though recognition has entered what appears to be a steady state.

Learning.—Proportion of correct responses to study-trials stimuli as a function of test-trial number are shown in Fig. 2, beginning with the test trial after the first study trial. Each of these proportions is based on (48 *S*s) × (6 stimuli per type) = 288 experimental events. Correct responding for the two pair types is seen in Fig. 2 to take the form of diverging functions: Whereas there is no difference in the early trials, over Trials 9–12 the superiority of Type-C pairs over Type-R pairs is highly significant statistically ($z = 5.79$, $p < .001$).

TABLE 1

Proportion Correct Responses (CR) Given Recognition (R) and Nonrecognition (NR) of the Stimulus

Statistic	Pair Type	
	C	R
P(CR/R)	.53	.44
P(CR/NR)	.31	.37
P(Chance CR)	.33	.33

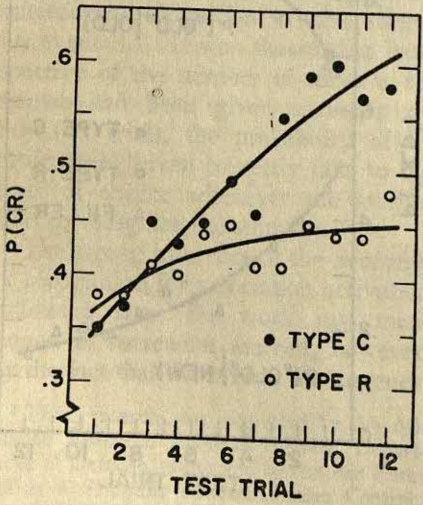


Fig. 2. Proportion correct responses (CR) to study-trials stimuli for Types C and R pairs. (The continuous curves are visual approximations.)

Note that over the 12 test trials there is a clear increase in correct responding for Type-R pairs. This result reflects an increase in *S*'s capability to retain the Type-R pairings current on a given study trial until tested a few seconds later.

Learning-recognition contingency.—Consider first the bottom-row entries in Table 1. The probability of a correct response by chance alone for each of the two pair types was arbitrarily set at .33. This was done on the basis of there being three response alternatives. As it turned out, proportion "correct" responses on the preliminary test trial prior to the first study trial was precisely .33 for the two pair types combined. When the two pair types are considered separately, proportion "correct" responses on this preliminary test trial was .31 for Type-C pairs, .36 or Type-R pairs.[2] Neither of these proportions differs significantly from the .33 a priori expectation. Over the two pair types together, the numbers, 1, 2, and 3, were chosen 33, 36, and 31% of the time on this initial test trial, indicating no response bias of any particular interest.

---

[2] These observed proportions were determined by defining "correct" to mean the pairing that was to obtain on the first study trial, which followed the preliminary test trial immediately.

The top two rows of Table 1 pertain to the learning-recognition contingency itself. Consider first the case for Type-C pairs. Out of (48 *S*s) × (6 stimuli) × (12 trials) = 3,456 test-trial events, there were 2,864 stimulus recognitions. Of these, 1,514, or 53%, resulted in a correct response. Out of the 592 nonrecognitions, only 182, or 31%, resulted in a correct response. The former percentage differs significantly from the chance rate of 33% ($z = 22.5$, $p < .001$), while the latter does not ($z = 1.05$, $p = .29$). Further, the percentage of correct responses given recognition significantly exceeds the percentage of correct responses given nonrecognition ($z = 9.8$, $p < .001$).

Consider now the case for Type-R pairs. Out of the 2.853 recognitions, 1,245, or 44%, resulted in a correct response. This percentage differs significantly from the chance rate of 33% ($z = 12.2$, $p < .001$). Out of the 603 nonrecognitions, 223, or 37%, resulted in a correct response. This percentage also differs significantly from the chance rate ($z = 2.0$, $p = .05$). Further, as with Type-C pairs, the percentage of correct responses given recognition of the stimulus significantly exceeds the percentage of correct responses given nonrecognition ($z = 3.0$, $p = .003$).

When comparisons between pair types are made, it turns out that proportion of correct recalls given stimulus recognition for Type-C pairs is significantly larger than that proportion for Type-R pairs ($z = 6.98$, $p < .001$). This relation, as can be seen in Table 1, is reversed for proportion correct recalls given nonrecognition of the stimulus ($z = 2.28$, $p = .02$).

From a joint consideration of Fig. 2 and Table 1, one can easily glean that proportion of correct responses given stimulus recognition, P(CR/R), increases with trials. This increase is regular, and especially marked for Type-C pairs. Over Trials 9–12: for Type-R pairs, P(CR/R) = .46; while for Type-C pairs, P(CR/R) = .62. The picture for proportion correct responses given nonrecognition of the stimulus, P(CR/NR), is not as regular. This statistic, for Trial Blocks 1–4, 5–8, and 9–12 is .35, .35, and .42, respectively, for Type-R pairs. The corresponding proportions for Type-C pairs are .25, .34. and .36. Thus although P(CR/NR) falls within the neighborhood of chance expectation, there are deviations. However, only the values .42 and .25 depart significantly from .33 ($z = 2.4$, $p = .02$, and $z = 2.9$, $p < .01$, respectively).

DISCUSSION

The picture that emerges is the following: Given stimulus recognition, correct responding occurs to the extent that association formation has progressed; given nonrecognition of the stimulus, correct responding is much reduced toward the chance level. Concerning this last, correct responding given nonrecognition was indistinguishable from chance for Type-C pairs, somewhat above chance for Type-R pairs.

Regarding Type-C pairs, the present finding that correct responding significantly exceeds the chance level given stimulus recognition but approximates the chance level given nonrecognition is in perfect agreement with earlier work (Martin, in press), where all pairs were Type C. Thus the original result is replicated.

Regarding Type-R pairs, two phenomena demand attention. First, correct responding given stimulus recognition significantly exceeds chance expectation. This result means that *S*s were capable of retaining, albeit to a limited extent, the new pairings presented on a given study trial until tested on the immediately following test trial. By chance alone, one might expect two correct responses out of six; however, by Trials 9–12, *S*s were emitting 2.8 correct responses out of a possible 6. This phenomenon may be identified as an instance of intratrial retention, the reality of which has been carefully scrutinized by Tulving (1964).

There are two plausible reasons why intratrial retention of Type-R pairs might increase with trials. (*a*) Owing to the gradual acquisition of the consistently paired Type-C pairs, the functional length of the entire list of 12 pairs is effectively shortened. Thus with mastery of an increasing number of Type-C pairs, more of the immediate-memory capacity of *S* can be assigned to retention of current Type-R pairings. (*b*) As practice proceeds, *S* becomes more familiar with the stimuli (CCCs), thereby depreciating stimulus indiscriminability as a source of confusion and/or interference.

The second phenomenon associated with Type-R pairs that requires comment is the finding that proportion of correct responses given nonrecognition of the stimulus significantly exceeds not only the chance level, but also the corresponding proportion for Type-C pairs. The present writer sees no explanation, either obvious or subtle, for this result. Because pair type was a within-*S*s factor, and because the Type-C, Type-R, and filler trigrams were presented in a new random order on every test trial, exercise of a differential "set" for Type-R pairs does not seem likely. Moreover, this result cannot be traced to differential peculiarities of the trigrams assigned to Types C and R, for each trigram occurred equally often over the six lists as a Type C and as a Type R. Finally, because from trial to trial the response assigned to a given Type-R stimulus was unpredictable, in contrast with the situation for Type-C pairs, it occurs to this writer that the result under consideration, namely, that correct responding given nonrecognition of the stimulus is superior for Type-R pairs relative to Type-C pairs, is a sampling error.

The foregoing notwithstanding, the results for Type-R pairs confirm at least the weaker hypothesis that stimulus nonrecognition interferes with association activation; the difference between proportion correct responding given stimulus recognition and proportion given nonrecognition is statistically significant and in directional agreement with the results for Type-C pairs.

The foregoing comments pertain to the first question posed at the outset, "What role does stimulus recognition play in paired-associate learning?" The second question was, "Does association of a response with a stimulus affect the recognizability of that stimulus?" In answer, three points need to be made. (*a*) Intratrial retention was sufficient to support above-chance correct responding, as evidenced by the learning data for Type-R pairs (Fig. 2). This means that *S*s were not attaching single associations to Type-R stimuli and exercising such associations without regard for new pairings as they occurred. Rather, they were attending to the new pairings and attempting to retain them until tested, thereupon dropping them in favor of the next pairing. Thus *S*s were in fact responding to Type-R stimuli with variable associations. (*b*) These variable (Type-R) associations were not of the same strength as the consistent (Type-C) associations (Fig. 2 and the top row of Table 1). This is a completely unsurprising result and merely indicates that associations can be strengthened with repetition. But, (*c*) stimulus recognition did not differ for Type-C and Type-R pairs, either at the outset, when correct responding for the two types was equal, or later in learning, when correct responding was clearly superior for Type-C pairs (Fig. 1). Thus the answer to the second question must be in the negative: the presence of strong, stable associations does not enhance recognition of stimuli, at least not in the experimental paired-associate setting.

REFERENCES

MARTIN, E. Stimulus recognition in aural paired-associate learning. *J. verbal Learn. verbal Behav.*, in press.

TULVING, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychol. Rev.*, 1964, **71**, 219–237.

WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received July 29, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 506–510

INCUBATION OF ANXIETY: EFFECT ON GENERALIZATION GRADIENTS[1]

OTELLO DESIDERATO AND MERTON E. WASSARMAN

Connecticut College

The Ss were trained to label a visual stimulus and then were tested —either immediately or after a delay—for generalization of the verbal response. ½ the Ss within each delay group received initial training under aversive, the other ½ under nonaversive, conditions. Each subgroup was further subdivided into high and low anxious Ss. Gradients of higher elevation were produced by aversive training and by anxious Ss. If anxious Ss were tested after a delay, generalized responses increased; if nonanxious Ss were delayed before testing, generalized responses declined. No effects on gradient slopes were found. Considered together with self-rating data, the results seemed to support an "incubation" of anxiety interpretation, provided certain assumptions were made regarding anxiety-arousal conditions among anxious Ss.

Evidence for an increase in autonomic reactivity with time elapsing between conditioning and testing (incubation effect) has been reported for human *S*s (Golin, 1961; Mednick, 1957). The main purpose of the present experiment was to reexamine the incubation effect in humans, by studythe effect of delayed testing on the generalization of a verbal response following training under aversive or nonaversive conditions. Support for the incubation hypothesis would take the form of a more elevated gradient when delayed testing followed training under aversive conditions. For *S*s trained under "neutral" conditions, there would be no basis for an anxiety incubation process and, hence, delay in testing should not affect gradient characteristics. On the other hand, extrapolation from animal studies (Desiderato, Butler, & Meyer, 1966; McAllister & McAllister, 1963) leads to the prediction that delayed testing of human *S*s would merely flatten the generalization gradient slope, regardless of the aversive nature of the training conditions.

In order to determine whether amount of incubation also varied with anxiety level, *S*s were selected on the basis of extreme scores on Heineman's (1953) forced-choice form of the Taylor Manifest Anxiety scale (*MA*s).

METHOD

The *S*s were 56 freshmen college women, drawn from the extreme 10% of a distribution of 330 anxiety-scale scores. Means for the high anxiety and low anxiety groups were 70.2 and 40.6, respectively.

During the training phase, a 104-mm.^{2} black square, appearing in the center of a circular white field 378 mm. in diameter, was projected 29 times on a screen placed 5½ ft. from *S*. Each presentation of the 104-mm. square (CS) lasted 3 sec., and a 5-sec. blank stimulus separated the trials. Each *S* was instructed to call out the number "4" immediately upon the termination of the CS. The verbal response was monitored via an intercom leading to an adjacent one-way vision room, from which *E* operated the apparatus and recorded responses.

On a randomly selected 83% of the training trials, an auditory stimulus (UCS) was delivered to *S* via Grason-Stadler earphones, for a 1.5-sec. duration beginning 1.5 sec. after CS onset. In each anxiety group, for half the *S*s the stimulus consisted of intense

---

[1] The authors thank Robert L. Rhyne for several statistical suggestions and David Liberman for assistance in constructing the apparatus.

white noise generated by a Foringer noise generator, Model 1291, at an intensity level of 119-db. SPL. For the remaining *S*s the white noise level was maintained at an innocuous level of 61-db. SPL. These procedures were designed to permit simultaneous observation of both verbal and PGR responses, following Eriksen's (1965) method. Although apparatus failure prior to the experiment made it impossible to include recording of autonomic reactivity, the procedure described was retained unchanged.

After the twenty-ninth CS trial, *S*s were tested for generalization of the verbal response to stimulus squares of different sizes. Each *S* was shown successively two black squares, each in the center of a circular white field, one 80 mm.^{2}, the other, 128 mm.^{2}, and was told that the stimulus she had been labeling "4" was actually the middle square in a series of seven squares, the largest and smallest being the 80- and 128-mm. squares. She was then instructed to carefully observe each of a series of squares to be projected on the screen, to wait until each one terminated, and to call out "4" after each square she judged to be identical in size to the one used on the training trials, saying "no" to every other stimulus.

The seven stimuli used in testing ranged from 80 to 128 mm.^{2}, in 8-mm. steps. Each stimulus occurred once in a block of seven trials, within which order was randomized except for the restriction that no two adjacent sizes occurred successively. The same order was maintained for all *S*s. A total of three blocks of trials was given, yielding three presentations of each stimulus size. As in the training phase, each stimulus appeared for a 3-sec. duration. Earphones were worn during all test trials, but no auditory signals were delivered.

For half the *S*s, the testing procedure followed the training phase immediately. For the remaining *S*s, testing was delayed for 15 min. Delayed *S*s were told they could leave the test room but not talk to anyone, and that after 15 min. the second part of the experiment would commence. The design thus consisted of high and low anxiety groups (HA vs. LA), each divided into aversive and nonaversive training conditions (A vs. NA), with each training subgroup further subdivided into groups receiving immediate or delayed testing (I vs. D).

In an attempt to obtain some indirect measure of differences in emotionality experienced by the various groups, at the end of the experiment every *S* was asked to respond to the following two rating scales:

1. During the first part of the experiment when all the squares were the same size, how did you feel whenever a projected square would signal the coming of a burst of noise?
   Circle one number: 1 2 3 4 5
   It didn't bother me at all — I was terribly upset
2. At the start of the second part of the experiment, did you feel more nervous or apprehensive than before?
   Circle one number: 1 2 3 4 5
   No change — I was much more nervous than before

RESULTS

The total number of responses to each stimulus, summed across all *S*s, is shown in Fig. 1. The linear component of the downward trend was significant for the three smallest as well as the three largest stimuli, *F* (1, 110) = 81.00 and 32.19, respectively, $p < .001$ for both. The quadratic component was significant in neither case. Thus, the procedures used were successful in producing bidirectional falling gradients.

For the main analysis of variance, the mean number of responses given

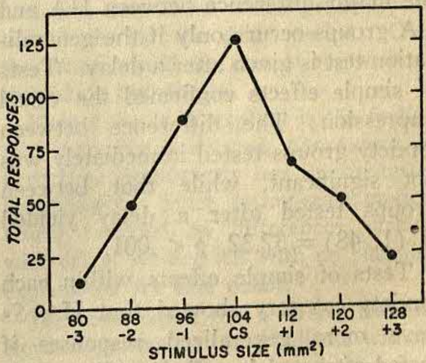


FIG. 1. Generalization gradient of "4" responses to stimuli greater and smaller than the 104 mm.^{2} CS. (Values include scores for all *S*s.)

by each *S* to stimuli ± 1 step from the CS was computed. The procedure was repeated for responses to stimuli ± 2 and ± 3 steps from the CS. Each mean score was then converted to $\sqrt{X + .5}$ in order to stabilize the variance. Thus, each *S* contributed one score to each of the three stimulus "distances" from the CS. Analysis of variance revealed a significant main effect for test stimuli, $F\ (2, 96) = 66.95$, $p < .001$, again indicating a pronounced generalization effect. The analysis also indicated that the group trained under aversive conditions (Group A) gave significantly more generalized responses than Group NA, $F\ (1, 48) = 12.52$, $p < .01$. However, all interaction effects which included the training variable failed to yield significant *F* ratios, indicating that aversiveness of the training conditions neither affected gradient slopes nor interacted with other variables to influence gradient elevation.

While HA *S*s gave significantly more generalized responses than the LA group, $F\ (1, 48) = 20.44$, $p < .001$, greater interest lies in the single significant interaction effect yielded by the analysis, the Anxiety × Delay interaction, $F\ (1, 48) = 12.39$, $p < .01$. As shown in Fig. 2, it is apparent that the major difference between HA and LA groups occurs only if the generalization test is given after a delay. Tests of simple effects confirmed the visual impression: The difference between anxiety groups tested immediately was not significant, while that between groups tested after a delay yielded $F\ (1, 48) = 32.32$, $p < .001$.

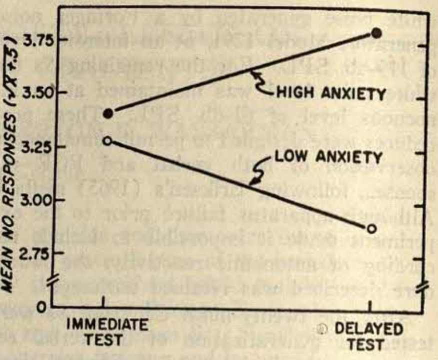


FIG. 2. Mean number of generalized responses given by high and low anxiety groups given immediate or delayed testing. (Values are based on transformed scores.)

Tests of simple effects within each anxiety category showed that HA *S*s gave more generalized responses if tested after a delay, $F\ (1, 48) = 5.98$, $p < .05$, a result compatible with an incubation notion. An unexpected finding was that LA *S*s gave significantly *fewer* generalized responses if testing was delayed, $F\ (1, 48) = 6.42$, $p < .05$. None of the other interactions was significant.

To determine whether self-reports of anxiety might correspond in some way to the generalization test findings, rating-scale data were analyzed using Mann-Whitney *U* tests. On the first item, which dealt with the experienced effect of the auditory stimulus used as the UCS in training, mean ratings of 2.93 and 1.29 were recorded by Groups A and NA, respectively. The mean difference is significant ($z = 5.66$, $p < .001$).[2] Further breakdown of Groups A and NA into anxiety subgroups (Table 1) revealed that a significant difference in experienced emotionality was shown by HA and LA *S*s if the aversive stimulus was used ($z = 2.27$, $p < .05$), but not if the auditory stimulus was nonaversive.

The second rating-scale item was intended to measure anxiety at the start of generalization testing. Since the Anxiety × Delay interaction in the analysis of generalized responses was significant, mean ratings for anxiety

---

[2] *P* values for all Mann-Whitney tests in this report are for two-tailed tests.

TABLE 1

Mean Anxiety Ratings of Several Groups Obtained Post-experimentally

Item	High Anxiety	Low Anxiety
Anxiety during Training		
Aversive UCS	3.36	2.50
Nonaversive UCS	1.43	1.14
Anxiety at Beginning of Testing		
Immediate Testing	3.29	1.50
Delayed Testing	4.36	1.50

groups tested immediately and after delay are presented in Table 1. While both HA subgroups indicated increased "nervousness" at the start of testing, a Mann-Whitney *U* test showed that the mean of the delayed subgroup was significantly greater ($z = 2.07$, $p < .05$). Inspection indicates that among LA *S*s, delay produced no effect at all.

Discussion

According to the incubation hypothesis, it was expected that delayed testing following aversive training would produce elevated generalization gradients, but would not affect their slope. This hypothesis predicts no gradient changes associated with delay, if the original training conditions are nonaversive. The failure to obtain a significant Aversiveness × Delay interaction does not support the notion of a simple increase in conditioned emotionality across time. Neither was support found for the more limited prediction which would restrict the incubation effect to anxious *S*s, since the Aversiveness × Delay × Anxiety interaction also was not significant.

The simple flattening of generalization gradients across time, previously reported for animal *S*s in both aversive (Desiderato, Butler, & Meyer, 1966; McAllister & McAllister, 1963) and appetitive (Perkins & Weyant, 1958; Thomas & Lopez, 1962) conditioning, was not found here. The nonsignificant Delay × Test Stimuli interaction also fails to corroborate recent findings for human *S*s. Using a temporal generalization method, Desiderato, Foldes, and Gockley (1966) reported that delayed testing merely reduced gradient slopes. Similarly, in a psychophysical study, increasing the time elapsing between the presentation of constant and variable stimuli has been shown to impair discrimination in humans (McAllister, McAllister, & Franchina, 1965).

One major result of the present study is that anxious *S*s will produce more elevated, but not steeper, generalization gradients if testing is delayed than if it proceeds immediately after training, regardless of the aversiveness of the training conditions. This result may be compatible with an incubation notion if it is assumed that, while a time delay may produce an increase in emotionality conditioned during aversive training, the simple requirement that the already-anxious *S* wait for a "second part of the experiment" to begin may be sufficient to cause a further increase in anxiety. In support of this view, further breakdown of responses given by anxious *S*s to the rating-scale item concerning anxiety at the *start* of testing showed greater anxiety among delayed than immediately tested groups, following *nonaversive* as well as aversive training. What is not clear, however, is whether the increased anxiety rating is the immediate result of instructions to wait or whether it measures uneasiness which gradually increases during the delay period. Only the latter would be compatible with an incubation hypothesis.

More difficult to explain is the finding that if testing is delayed, nonanxious *S*s generalize significantly less than if tested immediately. Since rating-scale anxiety scores revealed no differences varying with delay, we are left with no plausible explanation for this result.

The observation that aversive training elevated generalization gradients without influencing their slope extends the findings of temporal generalization studies employing human *S*s (Desiderato, 1964; Wenar, 1954). It may also be noted that the failure of the aversiveness variable to interact with anxiety level does not lend support to the "reactive" hypothesis which maintains that *MA*s scores reflect *potentiality* for emotional arousal in the presence of stressful conditions (Taylor, 1956).

REFERENCES

Desiderato, O. Effect of anxiety and stress on reaction time and temporal generalization. *Psychol. Rep.*, 1964, **14**, 51–58.

Desiderato, O., Butler, B., & Meyer, C. Changes in fear generalization gradients as a function of delayed testing. *J. exp. Psychol.*, 1966, **72**, 678–682.

Desiderato, O., Foldes, J. M., & Gockley, J. S. Incubation of anxiety: Effect on temporal generalization. *Psychon. Sci.*, 1966, **6**, 139–140.

Eriksen, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, **69**, 625–634.

Golin, S. Incubation effect: Role of awareness in an immediate vs. delayed test of conditioned emotionality. *J. abnorm. soc. Psychol.*, 1961, **63**, 534–539.

Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.

McAllister, W. R., & McAllister, D. E. Increase over time in the stimulus generalization of acquired fear. *J. exp. Psychol.*, 1963, **65**, 576–582.

McAllister, W. R., McAllister, D. E., & Franchina, J. J. Dependence of equality judgments upon the temporal interval between stimulus presentations. *J. exp. Psychol.*, 1965, **70**, 602–605.

Mednick, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55**, 315–321.

Perkins, C. C., Jr., & Weyant, R. G. The interval between training and test trials as a determiner of the slope of generalization gradients. *J. comp. physiol. Psychol.*, 1958, **51**, 596–600.

Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

Thomas, D. R., & Lopez, L. J. The effects of delayed testing on generalization slope. *J. comp. physiol. Psychol.*, 1962, **55**, 541–544.

Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

(Received July 29, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 511–516

S-R, R-S INDEPENDENCE AND THE INTERFERENCE POTENCY OF LATENT R-S ASSOCIATIONS

WILLIAM A. JOHNSTON

Ohio State University

8 groups of 16 Ss each were exposed to different 4-stage transfer paradigms in an assessment of associative symmetry: E-C, A-B, C-A, A-B; E-C, A-B, C-A, B-A; E-C, A-B, C-D, A-B; E-C, A-B, C-D, B-A; E-F, A-B, C-A, A-B; E-F, A-B, C-A, B-A; E-F, A-B, C-D, A-B; E-F, A-B, C-D, B-A. A stimulus in 1 list (e.g., A-B) was made "available" by having it serve a response role in another list (e.g., C-A). It was found that (a) R-S is formed during S-R learning, (b) R-S becomes manifest if the stimulus becomes available even after S-R learning, (c) latent and manifest R-S associations (e.g., A-C) can interfere retroactively with S-R (e.g., A-B) recall, and (d) S-R and R-S are independent such that interference of A-B does not affect B-A.

The "principle" of associative symmetry asserts that as S-R is learned, an equipotent R-S association is also formed (Asch & Ebenholtz, 1962). The accumulated evidence suggests that associative symmetry is at least approximated when the stimuli and responses are equally *available,* i.e., equally recallable (Ekstrand, 1966a; Horowitz, Norman, & Day, 1966). The apparent unidirectionality that arises in normal paired-associate learning stems in part from the fact that the response terms are generally more available than the stimulus terms (Newman, Cunningham, & Gray, 1965; Newman & Gray, 1964).

However, at least two questions remain unanswered. First, are the S-R and R-S associations independent, or are they simply different manifestations of a single, underlying association? Considerable dependence of S-R and R-S has been observed in retroactive inhibition (RI) experiments (Birnbaum, 1960; Greenbloom & Kimble, 1965; Houston, 1964), while evidence favoring independence has been found in studies of bidirectional learning (Battig & Koppenaal, 1965; Schild & Battig, 1966; Underwood & Keppel, 1963; Voss, 1965). Second, will the R-S association be exposed if the stimulus is not available during S-R learning but becomes available thereafter? That is, will a *latent* R-S association become *manifest* when the stimulus becomes available? Though Horowitz et al. (1966) speculate that the answer is positive and Ekstrand (1966a) implies that it is negative, a direct empirical test has yet to be conducted. Finally, if latent R-S associations exist, can they interfere with other associations? The present study was designed to shed further light on all of these issues.

METHOD

Design.—Horowitz et al. (1966) reported that the availability of an item ". . . grows fastest if the subject produces the item from memory [p. 3]." Thus, the stimulus of one pair (e.g., A-B) can be made available by having it serve as a response in another pair (e.g., C-A). This technique for producing stimulus availability was employed in the

---

[1] The writer is grateful to Judith Ryan, Richard Edison, James Jones, and Robert Yohman for help in collecting and analyzing the data.

present study. Each group learned four consecutive lists denoted first prior learning (PL-1), second prior learning (PL-2), interpolated learning (IL), and relearning (RL). The experimental design is shown in Table 1. Since PL-1 responses become IL stimuli only in Groups 1-4, the IL stimuli should be available only in Groups 1-4. Similarly, IL serves to make the PL-2 stimuli available in Groups 1, 2, 5, and 6 but not in the remaining groups. Finally, the PL-2 S-R associations are relearned during RL in the odd-numbered groups, while the PL-2 R-S associations are relearned in the even-numbered groups. Thus, a 2 (PL-1) × 2 (IL) × 2 (RL) factorial design was used.

Since the IL stimuli are made available in Groups 1–4, the IL R-S association should be manifest. This manifest IL R-S association should differentially affect RL performance in Groups 1–4. In Group 1 the IL R-S association is A-C, which association should produce RI of A-B, thereby retarding RL performance. By contrast, in Group 3 the IL R-S association is D-C, which association should interfere less with A-B relearning. Thus, Group 3 should surpass Group 1 at RL. Furthermore, if the A-B and B-A associations are interdependent, RI of RL should also occur in Group 2, and Group 4 should surpass Group 2 at RL. However, if A-B and B-A are independent associations, and if B-A becomes manifest when A becomes available by serving as a response during IL, then retroactive facilitation should occur in Group 2, and Group 2 should surpass Group 4.

Since the IL stimuli are not available in Groups 5–8, the R-S associations (if formed at all) should be latent. Thus, the design provided a test of the relative RI potency of latent (Groups 5–8) and manifest (Groups 1–4) R-S associations potentially formed during IL. For example, if a latent A-C association produces less RI than a manifest A-C association, then RI in Group 5 should be less than that in Group 1.

TABLE 1

EXPERIMENTAL DESIGN

Group	PL-1	PL-2	IL	RL
1	E-C	A-B	C-A	A-B
2	E-C	A-B	C-A	B-A
3	E-C	A-B	C-D	A-B
4	E-C	A-B	C-D	B-A
5	E-F	A-B	C-A	A-B
6	E-F	A-B	C-A	B-A
7	E-F	A-B	C-D	A-B
8	E-F	A-B	C-D	B-A

Note.—The Letters A through F represent unrelated stimulus and response terms.

Lists.—Each list was composed of six pairs so that 36 items were required to form the A, B, C, D, E, and F terms. The six E terms were single digits (1–6), and the 30 remaining items were obtained from Noble's (1952) list of disyllables and had *m* values ranging from .99 (GOJEY) to 2.63 (TARTAN). An effort was made to keep similarity between disyllables minimal while maintaining low *m* values. Low *m* values assured low initial availability of the disyllables. Digits were used as PL-1 stimuli to minimize proactive interference of IL S-R associations (C-A and C-D) stemming from PL-1 R-S associations (C-E) in Groups 1–4. The use of different item classes (e.g., digits vs. disyllables) is known to reduce interlist interference (Postman, Keppel, & Stark, 1965).

Two versions of each list were generated by two different drawings from the 36-item pool. Each drawing was random without replacement with the restrictions that digits serve only as E items, that there be six disyllables each for the A, B, C, D, and F terms, that disyllables in the same pair not begin with the same letter, and that the same disyllable not serve in the same list (e.g., PL-1) in the two versions. Five random orders of the six pairs were constructed for each list to reduce the opportunity for serial learning, and the five orders were arranged in a random sequence on a tape. The A-B pairs were ordered and sequenced independently for PL-2 and RL. Thus, the two versions each of the E-C, E-F, A-B (PL-2), C-A, C-D, A-B (RL), and B-A lists yielded 14 tapes.

Subjects.—The *S*s were 128 female undergraduates, some of whom had served in one previous paired-associate learning experiment in which nonsense syllable triads served as stimuli and digits served as responses. Sixteen conditions were formed from the factorial combination of the eight groups with the two list versions, and *S*s were assigned to conditions randomly with the restriction that there be eight *S*s per condition (16 *S*s per group).

Procedure.—The lists were presented at a 2:2-sec. rate on a Stowe memory drum with a 4-sec. intertrial interval and a 4-min. interlist interval. All lists were learned for 10 trials by the standard anticipation method. An 8-sec. interval was required between the fifth and sixth trials of each list so that the tape could be recycled. On the first trial for PL-1, PL-2, and IL the shutter was made inactive so that the stimulus and response members of each pair appeared together. Each pair was exposed for 8 sec. and *S* was required to pronounce the response term. If *S* mispronounced a response on one of these initial trials, *E* corrected her. The RL list was presented from the first trial through the tenth at a 2:2-sec. rate.

Standard paired-associate learning instructions were given at the start of the session, and *S*s were told that they were to relearn (A-B or B-A) immediately prior to the first RL trial. There follows a descriptive portion of the instructions for B-A RL: "The next list is one that you were presented before . . . The difference is that the responses will appear first and we want you to try to recall the original stimuli. Thus, the old associations have just been turned around."

RESULTS

Analyses of variance of total correct during PL-1, PL-2, and IL indicated comparable performance in the eight groups; F (7, 120) < 2.00 for each list. The mean number of correct responses on the last PL-2 trial was used to assess level of original A-B learning in the eight groups. These means were 5.19, 3.50, 4.06, 4.31, 4.56, 4.31, 4.31, and 4.13 for Groups 1–8, respectively, and an analysis of variance showed that they were ostensibly equivalent, F (7, 120) = 1.55, $p > .05$. Thus, level of A-B learning was nearly constant across groups. Nevertheless, loss and gain scores were used to assess performance differences between groups at RL since the use of these scores minimizes possible contaminating effects of even slight differences in A-B learning.

A loss (or gain) score was computed as the difference between the number of correct responses on the last PL-2 trial and the number correct on an RL trial for a given *S*. Superior performance on an RL trial denotes a gain, inferior performance denotes a loss. These data are summarized in Fig. 1. A Trials × PL-1 × IL × RL analysis of variance revealed trials as the only significant main effect, F (9, 1080) = 318.97, $p < .001$. However, the following interactions were also significant: IL × RL, F (1, 120) = 38.32, $p < .001$; Trials × IL × RL, F (9, 1080) = 7.57, $p < .001$; Trials × PL-1 × IL, F (9, 1080) = 2.43, $p < .05$; Trials × PL-1 × RL, F (9, 1080) = 2.49, $p < .05$; and Trials × PL-1 × IL × RL, F (9, 1080) = 2.02, $p < .05$. Duncan's test (Edwards, 1963), using the 5% level of significance, was employed to specify the loci of these interactions. The bases of the first two significant interactions are evident in Fig. 1. Specifically, Groups 2, 3, 6, and 7 tended to surpass Groups 1, 4, 5, and 8, particularly on initial RL trials. Hence, B-A relearning was inferior to A-B relearning only when C-D was learned during IL. By contrast,

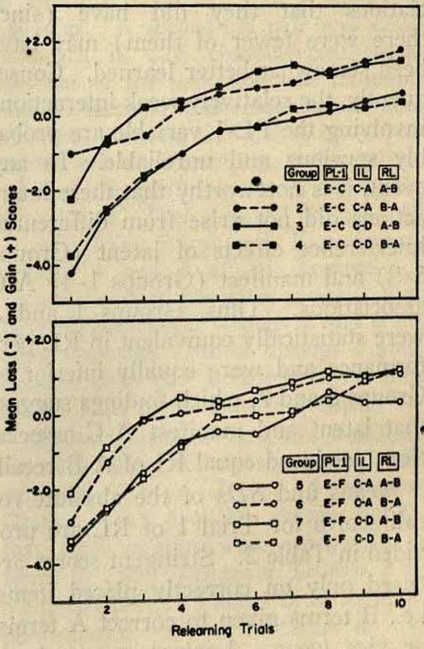


FIG. 1. Loss and gain scores on the RL trials for each group.

when C-A was acquired during IL, B-A relearning actually surpassed A-B relearning. Of course, all of these differences diminished across RL trials as performance approached an upper limit. These findings are highly critical as they indicate that B-A is formed during A-B acquisition, that B-A becomes manifest when A is made available after A-B learning, but that the two associations are independent such that the IL R-S association (A-C) retards A-B RL but not B-A RL.

The remaining significant interactions attained only borderline significance and stemmed from an unanticipated superiority of Group 2 over all other groups on the first RL trial. This superiority may be attributable to the fact that Group 2 was slightly, albeit not significantly, inferior to the other groups in terms of PL-2 learning. Compared to the *S*s of other groups, then, Group 2 *S*s had somewhat fewer PL-2 associations to lose and the associations that they did have (since there were fewer of them) may have been somewhat better learned. Consequently, the relatively weak interactions involving the PL-1 variable are probably spurious and unreliable. In any event, it is noteworthy that these interactions did not arise from differential interference effects of latent (Groups 5–8) and manifest (Groups 1–4) A-C associations. Thus, Groups 1 and 5 were statistically equivalent in RL performance and were equally inferior to Groups 3 and 7, which findings suggest that latent and manifest A-C associations produced equal RI of A-B recall.

TABLE 2

*M*s AND *SD*s OF LENIENT AND STRINGENT ABSOLUTE RECALL SCORES ON THE FIRST RL TRIAL

Group	PL-1	IL	RL	Lenient Score		Stringent Score	
				M	*SD*	*M*	*SD*
1	E-C	C-A	A-B	1.00	1.00	1.00	1.00
2	E-C	C-A	B-A	3.00	1.62	2.50	1.58
3	E-C	C-D	A-B	2.25	1.52	1.88	1.46
4	E-C	C-D	B-A	.56	.86	.56	.86
5	E-F	C-A	A-B	1.00	1.12	.94	.97
6	E-F	C-A	B-A	1.69	1.16	1.38	1.00
7	E-F	C-D	A-B	2.94	1.52	2.25	1.24
8	E-F	C-D	B-A	.88	.79	.69	.75

Means and *SD*s of the absolute recall scores for Trial 1 of RL are provided in Table 2. Stringent scores are based only on correctly placed items, i.e., B terms given to correct A terms, or vice versa. Lenient scores, however, include misplaced and mispronounced items as well. Therefore, stringent scores measure associative strength predominantly, while lenient scores tap item availability as well. It is important to note that the absolute recall scores of Table 2 are interrelated in the same way as are the loss scores of Fig. 1. Moreover, the near identity of the lenient and stringent scores suggests that the experimental variables operated primarily on associative strength. However, the fact that Groups 2, 3, 6, and 7 surpassed Groups 1, 4, 5, and 8 slightly more in terms of lenient scores than in terms of stringent scores indicates that the former groups had a slight advantage over the latter groups in terms of the availability of the items serving as responses as well as in terms of the strength of the RL associations. On the other hand, this difference between lenient and stringent scores may reflect a disparity in response *confidence* rather than a disparity in item availability (Ekstrand, 1966b).

Discussion

The present study furnishes further evidence that R-S is acquired during S-R learning and is exposed (made manifest) when the stimulus is available (Horowitz et al., 1966). However, the present data extend previous findings in showing that high stimulus availability is not prerequisite to the *formation* of an R-S association. That is, R-S is exposed even if the stimulus is made available *after* the termination of S-R learning (Groups 2 and 6). If the stimulus remains unavailable, however, the R-S association remains latent (Groups 4 and 8). These findings support the speculations of Horowitz et al. (1966).

Yet latent R-S associations are not necessarily dormant or inactive, for in Group 5 of the present study the latent R-S associations formed during IL (A-C) interfered with the S-R associations acquired during PL-2 (A-B). In fact, the amount of RI produced by the latent A-C associations in Group 5 was statistically equivalent to that produced by the manifest A-C associations in Group 1. This fact points to the potential importance of latent backward associations in governing overt behavior.

Furthermore, the present results indicate that S-R and R-S are independent. Specifically, the RI stemming from A-C reduced A-B recall (Groups 1 and 5) but not B-A recall (Groups 2 and 6). This finding is incompatible both with the notion of associative symmetry as stated by Asch and Ebenholtz (1962), and with the results of previous studies (Birnbaum, 1966; Greenbloom & Kimble, 1965; Houston, 1964). However, the present study tested differential RI of A-B and B-A using an A-B, C-A paradigm, while A-B, A-C (Birnbaum, 1966; Houston, 1964) and A-B, C-B (Greenbloom & Kimble, 1965) paradigms were employed in the previous studies. The use of different paradigms may account for the different results. Nonetheless, the present findings, along with the data on bidirectional learning (Battig & Koppenaal, 1965; Schild & Battig, 1966; Underwood & Keppel, 1963; Voss, 1965), bear strong testimony to the proposition that S-R and R-S are not *wholly* dependent. Clearly, therefore, a modification of the theory of associative symmetry as stated by Asch and Ebenholtz (1962) is in order.

As a whole, the present data support a four-stage analysis of learning in which S, R, S-R, and R-S learning stages are distinguished (Ekstrand, 1966a). This analysis may be especially helpful in understanding transfer phenomena. For example, the observed RI in Groups 1 and 5 of the present study indicates that the R-S stage of training can be an important determinant of S-R performance at transfer.

In short, then, the present results may be accounted for by assuming that B-A is formed during A-B learning but remains latent until A becomes available, that latent and manifest R-S associations have equal interference potency, and that A-B and B-A are independent associations. However, the recent findings of Young and Bickerstaff (1966) suggest an alternative account of the present data. These investigators had *S*s learn two consecutive paired-associate lists in which the first list was systematically reconstituted to form the second list. In no case were the same two items paired together in both lists. Interlist change in item function (CIF) occurred if a first-list stimulus was a second-list response, or if a first-list response was a second-list stimulus. A second-list pair was easiest to learn if CIF characterized neither member of the pair, and was most difficult to learn if CIF characterized both members. Hence, CIF per se has a detrimental effect on learning, and may have affected RL performance in the present experiment. Thus, Control Groups 3 and 7 may have surpassed Experimental Groups 1 and 5 at RL because CIF occurred only in the latter groups. Similarly Experimental Groups 2 and 6 may have surpassed Control Groups 4 and 8 because more CIF occurred in the latter groups than in the former groups.

However, several facts are incompatible with this CIF interpretation. For example, Groups 2 and 6 surpassed Groups 1 and 5 at RL despite the fact that these groups were comparable in terms of CIF. Moreover, Groups 2 and 6 equaled Groups 3 and 7 in terms of RL performance but not in terms of CIF. Finally, the intergroup differences in CIF at IL did not produce differences in IL performance. All of these facts indicate that CIF effects were minimal in the present study, and point to the inadequacy of the CIF account of the RL data. Still, CIF must be considered as a potential source of variance in studies of associative symmetry.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. The principle of associative symmetry. *Proc. Amer. Phil. Soc.*, 1962, **106**, 135–163.

Battig, W. F., & Koppenaal, R. J. Associative asymmetry in S-R vs. R-S recall of double function lists. *Psychol. Rep.*, 1965, **16**, 287–293.

Birnbaum, I. M. Unlearning in two directions. *J. exp. Psychol.*, 1966, **72**, 61–67.

Edwards, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1963.

Ekstrand, B. R. Backward associations. *Psychol. Bull.*, 1966, **65**, 50–64. (a)

Ekstrand, B. R. A note on measuring response learning during paired-associate learning. *J. verbal Learn. verbal Behav.*, 1966, **5**, 344–347. (b)

Greenbloom, R., & Kimble, G. A. Extinction of R-S associations and performance on recall trials without informative feedback. *J. verbal Learn. verbal Behav.*, 1965, **4**, 341–347.

Horowitz, L. M., Norman, S. A., & Day, R. S. Availability and associative symmetry. *Psychol. Rev.*, 1966, **73**, 1–15.

Houston, J. P. Verbal R-S strength following S-R extinction. *Psychon. Sci.*, 1964, **1**, 173–174.

Newman, S. E., Cunningham, J. W., & Gray, C. W. Asymmetry in stimulus-term and response-term recall following paired-associate learning. *Psychon. Sci.*, 1965, **2**, 297–298.

Newman, S. E., & Gray, C. W. S-R vs. R-S recall and R-term vs. S-term recall following paired-associate training. *Amer. J. Psychol.*, 1964, **77**, 444–450.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

Postman, L., Keppel, G., & Stark, K. Unlearning as a function of the relationship between successive response classes. *J. exp. Psychol.*, 1965, **69**, 111–118.

Schild, M. E., & Battig, W. F. Directionality in paired-associate learning. *J. verbal Learn. verbal. Behav.*, 1966, **5**, 42–49.

Underwood, B. J., & Keppel, G. Bidirectional paired-associate learning. *Amer. J. Psychol.*, 1963, **76**, 470–474.

Voss, J. F. Effect of pairing directionality and anticipatory cue in paired-associate learning. *J. exp. Psychol.*, 1965, **69**, 490–495.

Young, R. K., & Bickerstaff, T. R. Change of item function in paired-associate learning. *J. exp. Psychol.*, 1966, **72**, 514–518.

(Received August 1, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 517–524

CHOICE REACTION TIME AS A FUNCTION OF STIMULUS UNCERTAINTY, RESPONSE UNCERTAINTY, AND BEHAVIORAL HYPOTHESES[1]

IRA H. BERNSTEIN, DONALD L. SCHURMAN,[2] AND GENE FORESTER

Arlington State College of the University of Texas

2 studies were concerned with the effects of stimulus and response uncertainty upon reaction time (RT). In Experiment I, it was found that changes in stimulus uncertainty produced linear changes in RT but changes in response uncertainty produced a step function with a maximum at 1 bit of response uncertainty. Correct pretrial guesses as to which stimulus event would occur tended to facilitate RT only in the more complex response-uncertainty conditions. Experiment II replicated the finding that stimulus uncertainty per se is a variable effecting RT. The results were discussed relative to the role of attention and response conflict.

Recent studies of choice reaction time have been concerned with the mechanism underlying this form of information processing. Although there is considerable evidence to suggest that one relevant consideration is the response-selection process (Dillon, 1965; LaBerge & Tweedy, 1964; Morin, Forrin, & Archer, 1961), the results are equivocal as regards stimulus selection since only the last cited of the above articles found evidence for the contribution of stimulus uncertainty, U(s), as opposed to response uncertainty, U(r).

A strategy which affords partial isolation of U(s) and U(r) involves varying the number of stimuli associated with each response, a design referred to as information reducing by Posner (1964). Differential comparison of RTs in conditions varying in U(s) for fixed U(r) and vice versa achieve a partial isolation of the roles of stimulus and response processes, respectively. The above cited studies all used variations upon this strategy. However, they differed as to the discriminanda; the first two required *S*s to discriminate meaningul symbols (numerals and letters) whereas the latter employed rectangles differing in color. That they were not in substantive agreement suggests one possibility, stimulus-selection time is a function of the discriminanda.

To the present author's knowledge, spatially arrayed lights have not been employed as discriminanda in any task other than an information-conserving one, in which each stimulus event is uniquely paired with a particular response. Hence, an empirical problem is raised as to whether or not variation in U(s) contributes to variation in RT for this type of discriminanda.

A corollary interest in this problem arose from the results of a previous study concerned with the role of pretrial guesses in the RT situation (Bernstein & Reese, 1965). In this latter

---

[1] This experiment was supported by Grant MH-11173-01, National Institute of Mental Health, to the senior author. A version of Exp. I, conducted by DLS as an undergraduate honors project, was presented at the 1966 Convention of the Midwestern Psychological Association and a version of Exp. II, conducted by GF, was presented at the 1966 Convention of the Psychonomic Society. The authors wish to acknowledge their appreciation to Joseph Lappin for his comments regarding the manuscript.

[2] Now at the University of Illinois, Urbana.

study, trained *S*s made pretrial guesses as to which stimulus event they expected to occur in conditions varying in transmitted information, U(s:r), even though they were informed of the random nature of the presentation sequence. The results of this study for all trials replicated the traditional linear relation between RT and U(s:r); however, when correct and incorrect guesses were separately analyzed, it was found that this relation was obtained only when *S*s' guesses were incorrect. Reaction time was independent of U(s:r) for correct pretrial guesses.

Discussion of these results centered around the relation of these guesses to the construct behavioral hypothesis (BH), an implicit self-instructional set indexed through verbal reports that is assumed to exert propositional control over subsequent behavior (Dulany, 1962). In the present context, the symbols BH+ and BH− will denote correct and incorrect anticipations of the forthcoming stimulus event, respectively. Because this study employed an information-conserving task, U(s), U(r), and U(s:r) were mutually confounded. Hence, the assumption that this facilitation occurred with respect to the response-selection process, assumed within the context of Dulany's theoretical approach to the study of set, could not be ascertained.

EXPERIMENT I

Experiment I was conducted as an extension of the Bernstein and Reese (1965) study involving the addition of various information-reducing conditions to the information-conserving conditions employed by them. The broad intent of Exp. I was an inquiry into the general nature of stimulus- and response-selection processes occurring to spatially arrayed lights and a further inquiry into the specific locus of self-instructional sets in RT.

Method

Subjects.—Five advanced psychology majors at Arlington State College, four female and one male, served as paid *S*s.

Apparatus.—A display panel inclined approximately 60° to horizontal was used as the stimulus panel. The stimuli were eight NE-51 neon lamps set in a 4 × 2 pattern on 3-in. lateral centers. An additional NE-51 lamp placed in the center of the panel served as a warning indicator.

The response panel was set directly below the stimulus panel and was inclined approximately 30° to horizontal. Eight push buttons were set at the corners of two 2½-in. square patterns. The squares were 3 in. apart. The push buttons were of n.o. type, approximately ⅝ in. in diameter and protruded ¼ in. above the surface of the panel. Both the lights and buttons were labeled L (left) and R (right) for the appropriate square of four. Each set of four was labeled one through four in left to right order.

The display unit was connected to the control unit by a 4-ft. cable. From the control unit, *E* could control selection of the stimulus presented, length of foreperiod delay (by means of a Hunter delay timer), and readout and reset of a Lafayette clock timer.

Procedure.—The *S*s were seated in a moderately well-lit display unit with their forefingers resting on an "x" in the center of each square of buttons. The foreperiod was a constant 2 sec. The warning light remained on throughout this time. The stimuli were presented in random order with single replacement in blocks of 16 trials. Various random orders were used to control learning effects, etc. The *S*s participated in simple reaction-time sessions until their initial practice effects had reached apparent asymptote as inferred from their RT distributions.

The *S*s were tested in each of 10 conditions following these practice sessions. Each condition was an information-reduction or information-conservation combination of U(s) and U(r) values of 0, 1, 2, and 3 bits/trial. Stimulus uncertainty was varied by *E*'s instructions eliminating seven, six, four, or none of the stimulus alternatives as a possible event for that trial. Response uncertainty was varied by the restrictions imposed by U(s) and an electronic switching circuit in the control unit which mapped eight, four, two, or one stimulus events onto a response button. The 10 conditions were run in different orders for each *S*, providing a partial counterbalancing. Presented in Table 1 is a summary of the conditions, the s-r mappings, the buttons controlling the lights, and the instructions for each condition. These conditions will be referred to by the value of U(s)-U(r) for that condition. Thus, the 2–1 condition refers to four possible stimulus events mapped onto two response buttons.

During each experimental session, *S*s were given 16 blocks of trials under the same condition, the first 4 of which were practice. Thus, a maximum of 192 RTs were obtained per session. Reaction times longer than .75 sec. were excluded from data analysis as these typically represented choice of incorrect buttons. Groups of four blocks were separated from each other by a 2-min. rest period. Sessions were separated by a minimum of 1 hr.

Results

Effects of U(s) and U(r).—The average RTs for the 10 experimental conditions are presented in Table 2, separately for each *S* and as a composite across *S*s. Separate simple analyses of variance were conducted for each *S*; differences across conditions were significant in all cases beyond the .001 level. Because of the necessary nonorthogonality of U(s) and U(r), a two-way analysis of variance could not be performed to evaluate the effects of these variables. Instead, a more indirect procedure was employed by analyzing differences through Tukey's gap, straggler, and excessive variability criteria (Edwards, 1958). The results of these tests are also presented in Table 2 where the conditions are rank ordered for each *S* such that groups receiving the same rank do not differ from other groups receiving this rank, but differ from groups assigned other ranks at the .01 level.

The results of Table 3 indicate that both U(s) and U(r) are related to RT. More explicitly, for a fixed value of U(s), conditions differing in U(r) are generally assigned to separate groups. For $U(s) = 1$, three of five *S*s show differences among levels of U(r). All five *S*s show such differences across levels of U(r) for $U(s) = 2$ bits and 3 bits. Conversely, the same trends are observed when different values of U(s) are examined for fixed levels of U(r). For $U(r) = 0$ and 2 bits, all *S*s show significant

TABLE 1

SUMMARY OF CONDITIONS, S-R MAPPINGS, BUTTONS CONTROLLING THE LIGHTS, AND INSTRUCTIONS FOR EACH CONDITION

Cond.	U(s)	U(r)	S-R Mapping	Controlling Button(s)	Instructions
0-0	0	0	1 to 1	each	I will tell you which light will come on and you repeat after me.
1-0	1	0	2 to 1	L1, L3 R1, R3	I will tell you which pair (e.g. L1 or L2, R3 or R4) and you guess which one.
1-1	1	1	1 to 1	each	(Same as Cond. 1-0)
2-0	2	0	4 to 1	L1, R1	I will tell you which side (R or L) and you guess which light.
2-1	2	1	2 to 1	L1, L3, R1, R3	(Same as Cond. 2-0)
2-2	2	2	1 to 1	each	(Same as Cond. 2-0)
3-0	3	0	8 to 1	L1 ½ time R1 ½ time	You guess which of the eight lights.
3-1	3	1	4 to 1	L1, R1	(Same as Cond. 3-0)
3-2	3	2	2 to 1	L1, L3, R1, R3	(Same as Cond. 3-0)
3-3	3	3	1 to 1	each	(Same as Cond. 3-0)

TABLE 2

MEAN RT IN MSEC. AS A FUNCTION OF U(S) AND U(R) AND RANKINGS OF EXPERIMENTAL CONDITIONS ACCORDING TO TUKEY'S CRITERIA (IN PARENTHESES)

S	0[a]	1[a]		2[a]			3[a]			
	0[b]	0[b]	1[b]	0[b]	1[b]	2[b]	0[b]	1[b]	2[b]	3[b]
TP	356(1)	378(2)	380(2)	398(2)	427(3)	424(3)	396(2)	431(3)	477(4)	493(4)
GR	315(1)	327(1)	369(3)	350(2)	394(3)	387(3)	381(3)	385(3)	423(4)	412(4)
AH	342(1)	357(2)	376(2)	387(2)	429(3)	402(3)	381(2)	433(4)	475(5)	479(5)
CT	282(1)	279(1)	329(2)	288(1)	346(4)	383(4)	316(2)	436(6)	405(5)	408(5)
EB	286(1)	319(2)	371(3)	319(2)	375(3)	372(3)	316(2)	429(4)	411(4)	429(4)
Composite	316	332	364	348	394	393	357	423	437	433

Note.—Conditions assigned different ranks differ at the .01 level.

[a] U(s).

[b] U(r).

differences among U(s) conditions, whereas for U(r) = 1 bit, four of five *S*s show significant differences.

Figures 1 and 2 portray these trends graphically as plots of the composite RT means as a function of U(r) for levels of U(s), and as a function of U(s) for levels of U(r), respectively. This composite reflects trends common to all *S*s. These graphs reflect an important difference between the U(r)-RT and U(s)-RT relationships. Changes in RT are linear with changes in U(s) with a slope that is a function of U(r). In contrast, changes in RT are essentially a step function of U(r) with the magnitude of the step determined by U(s).

The composite findings under the information conserving Cond. 0-0, 1-1, 2-2, and 3-3, presented in Fig. 3, are in accord with prior findings; the fit of these conditions to a straight line is good. Moreover, the slope parameter was 40 msec., a figure which closely agrees with the results of the initial

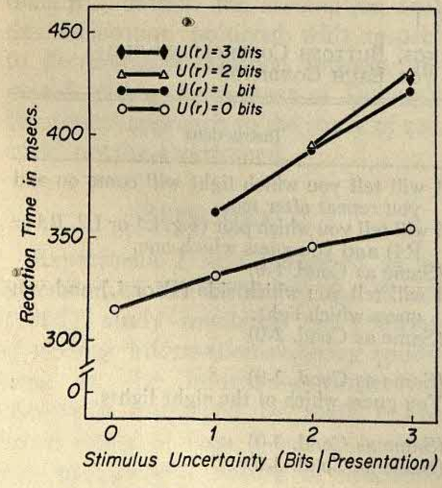


FIG. 1. RT as a function of U(s) for levels of U(r).

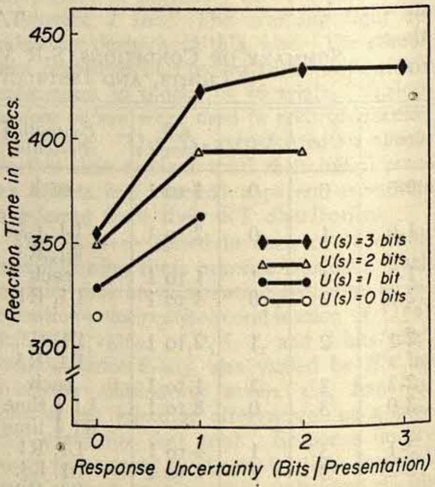


FIG. 2. RT as a function of U(r) for levels of U(s).

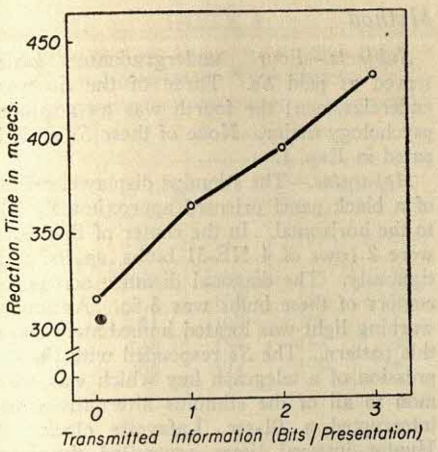


FIG. 3. RT at a function of U(s:r), information conservation conditions.

study in this series (Bernstein & Reese, 1965).

Effects of differences in BH.—Table 3 presents the mean difference score $(X_{BH+} - X_{BH-})$, the significance of these difference scores for each *S*, and the composite mean across *S*s for each experimental condition. Significance levels were established by means of *F* tests run within each condition and *S*. Reaction time differences were positive in 39 of the 45 comparisons and significant beyond the .05 level in 12 of these, confirming the facilitating effects found previously on BH+ trials. Although the overall significance levels are lower than those reported by Bernstein and Reese (1965), this was not unexpected since only half the number of observations were run per condition in the present study, as compared to the latter, due to the greater number of conditions to be run. Furthermore, the absolute magnitude of facilitation was comparable to that observed by Bernstein and Reese (1965).

In general, these results are not completely unequivocal. Generally, BH+ facilitation occurred with high levels of U(r) but one *S* (CT), showed a significant degree of facilitation in the 2-0 condition where there was no requisite-response selection. However, as analysis of the components of the Treatments × *S*s matrix indicated, the differences among difference scores are essentially due to a generally high level of BH+ facilitation that occurred when U(r) was ≥ 2 bits and a low level of facilitation that occurred when U(r) was < 2 bits.

The specific results of this analysis indicate that variability among the 3-0, 3-1, 3-2, and 3-3 conditions was significant beyond the .05 level, $F(3, 32) = 2.94$, and variability among the 2-0, 2-1, and 2-2 conditions approached significance, $F(2, 32) = 2.64$, $p < .10$.

TABLE 3

MEAN RT FACILITATION (BH+-BH−) AS A FUNCTION OF U(S) AND U(R)

S	1[a]		2[a]			3[a]			
	0[b]	1[b]	0[b]	1[b]	2[b]	0[b]	1[b]	2[b]	3[b]
TP	17	5	−20	33	59**	21	2	71**	1
GR	7	51**	23	36*	39**	27	−2	39	43*
AH	13	33*	16	18	18	19	3	14	41
CT	−3	26*	17**	43	81**	12	−1	56**	100**
EB	17	17	15	−16	17	15	35	−7	49
Composite	10.2	26.4	10.2	22.8	42.8	18.8	7.4	34.6	46.8

Note.—Values of p are based upon *F* tests run separately for each *S* in each condition.
[a] U(s).
[b] U(r).
* $p < .05$.
** $p < .01$.

Parallel comparisons of U(s) for fixed levels of U(r) were all nonsignificant ($F < 1.00$ in all cases).

Error rates.—Analysis of error rates indicated a range of 2–7% across *S*s with an overall average of about 4%. There was a trend towards increase in percentage of errors for the more complex U(s) and U(r) conditions but this was not significant.

EXPERIMENT II

Experiment II was conducted primarily as a partial replication of the U(s) differences observed in Exp. I. One possible interpretation of these findings is that differences in retinal sensitivity rather than U(s) may be causal since the lights were relatively well spaced about the fovea for U(s) = 3 bits. Poffenberger (in Woodworth & Schlosberg, 1955) has shown that RT is directly related to the distance from the retinal projection of the target to the fovea, a factor which provides empirical support for this possible argument. Although the results of Exp. I cannot be completely explained on these grounds since RT increased linearly in spite of a decrease in spatial disparity between U(s) values of 1 and 2 bits (in the former case, the 2 lights were 6 in. apart, in the latter the 4 lights were situated about the vertices of an imaginary 3-in. square), a replication was felt desirable for two other reasons. The first of these was the desire to see if effects could be obtained under conditions that a priori would appear least likely to generate them: All stimulus events were presented to the fovea, a simple telegraph key depression to any stimulus event was all that was required of *S*. Secondly, there was the desire to investigate the possible interaction of uncertainty imposed by varying the number of stimulus events and the utilization of a variable foreperiod delay.

Method

Subjects.—Four undergraduate males served as paid *S*s. Three of the *S*s were underclassmen; the fourth was an advanced psychology major. None of these *S*s participated in Exp. I.

Apparatus.—The stimulus display consisted of a black panel oriented approximately 30° to the horizontal. In the center of this panel were 2 rows of 4 NE-51 bulbs, spaced contiguously. The diagonal distance across the centers of these bulbs was 3 in. An amber warning light was located immediately above this pattern. The *S*s responded with the depression of a telegraph key which was common to all of the stimulus alternatives and interrupted a .01-sec. Lafayette clock. A Hunter interval timer controlled the foreperiod delay.

Procedure.—The *S*s were seated in a moderately well-lit room at a distance of 5 ft. from the display, with the response key located immediately under the index finger of their preferred hand. Each *S* participated in 6 sessions of approximately 1-hr. duration. The first session was denoted as practice and the results not analyzed. Following a set of practice trials, 48 RTs were obtained, per session, in each of four conditions, produced by the factorial combination of (*a*) fixed (1.5 sec.) or variable (1.0, 1.5, or 2.0 sec.) foreperiod delays and (*b*) 0 or 3 bits of U(s). This latter variable was manipulated, as in Exp. I, by instructing or not instructing *S* which of the 8 possible stimulus events would occur. Trials were grouped into blocks of 16 trials each in the same conditions. Randomization was done according to a Latin square principle such that each condition appeared once in a consecutive group or 4 blocks; different Latin squares were used for each session. BH— reports were obtained in both conditions for which U(s) = 3 bits.

Masking noise was generated by an electric fan enclosed in a cardboard box. One minor procedural difference between Exp. I and II was that trials exceeding 500 msec. were rerun to facilitate data analysis.

RESULTS

In Table 4 are presented the mean RTs as a joint function of U(s) and foreperiod variability, presented separately for each *S* and as a composite, pooled across sessions. Reaction times were analyzed by means of a three-way factorial format—sessions, 0 vs. 3 bits

TABLE 4

MEAN RT IN MSEC. AS A FUNCTION OF FOREPERIOD VARIABILITY

U(s) in bits	Foreperiod Delay			
	Fixed		Variable	
	0	3	0	3
S: RO	162	162	173	177
KL	169	177	184	194
RR	181	193	202	212
JP	148	154	158	166
$\bar{X}$	165	171	179	187

of U(s), and fixed vs. variable foreperiod delay—separately for each *S*.

The main effect of U(s) was significant beyond the .001 level for three of the four *S*s, replicating the desired result. The fourth *S* (RR), showed performance trends consistent with the remaining *S*s until the last session when his RTs in the 0-bit-fixed foreperiod delay conditions were inexplicably large. This reversal in performance reduced his overall *F* ratio for U(s) to nonsignificance (it was of borderline significance across the first four sessions) and produced a highly significant triple interaction, $p < .001$.

The remaining main effects were significant in all cases beyond the .001 level. Consistent across *S*s was a general decrease in RT across Sessions 1–4 followed by a slight rise in Session 5, probably due to motivational factors. In contrast, the interactions were in general nonsignificant; in no case was a given interaction significant beyond the .01 level for more than two *S*s. In other words, differences between 0 and 3 bit conditions were generally consistent across sessions and foreperiod delay conditions.

Additional analyses indicated that no differences between BH+ and BH− trials were significant; these differences ranged between ±10 msec. in no consistent pattern. Because some *S*s reported that the inner four lights in the pattern appeared more clearly in focus than the outer four, a comparison was made of the RTs between them. These comparisons were also negative.

DISCUSSION

The results of both studies clearly indicate that determinants of RTs to spatially arrayed lights include the independent contribution of stimulus and response uncertainty effects. In this respect, the present data clearly support the position of LaBerge and Tweedy (1964). In their study, *S*s were required to press one button when a green rectangle appeared and another button when either a red or blue rectangle appeared. Despite the response common to red and blue rectangles, differences in RTs appeared as a function of presentation probability, and hence U(s), to these two stimulus events. The present findings are an important extension, however, since no discrimination was required as to the specific stimulus event for correct-response selection in the U(r) = 0 bit conditions of Exp. I and II. In contrast, it was not possible for *S*s to gate the physical dimension along which U(s) was varied in the LaBerge and Tweedy (1964) study; stimulus selection was required prior to response selection.

These results suggest that the type of discriminanda effect the time required for stimulus selection. Stimulus processing time for familiar alpha-numeric symbols has been established to be of negligible duration (Brainard et al., 1962; Dillon, 1966; Morin, Forrin, & Archer, 1961) relative to that for spatially arrayed lights, given equally compatible s-r relationships. Although the mean time required for stimulus selection was only 10 msec. in Exp. II, this would appear to be a minimum rather than modal time estimate given the experimental contingencies employed. However, variables other than the type of discriminanda need to be considered. In particular, the systematic changes in the U(s)-RT function under the parameter U(r) resemble those results obtained in the visual recognition situation by Eriksen and Collins

(1965) and discussed under the term "attentional focusing." The similarity in these results lies in the dependence of the recognition process upon the level of requisite-response discrimination if it is assumed that U(r) is a measure of the degree of response discrimination.

The step function relating U(r) to RT appears of general interest to recent discussions concerned with the ability of a human *S* to process information and the particular question as to when increases in U(s:r) degrade RT. However, there appears to be one major interpretive problem in considering the factors producing this function. If U(s) is held constant and U(r) is increased, two simultaneous effects necessarily occur. The first of these is an increase in the number of responses which would be expected to lengthen RT through increased possibilities for response conflict. In contrast, a lower s-r mapping ratio is produced, which Posner (1964) has indicated tends to decrease RT. If it is assumed that these effects approximately offset one another, and if it is further assumed that this response-selection process is partially bypassed when U(r) is 0, then the observed results may be considered as largely the result of antagonistic and counteracting processes. Future research need be concerned with a disjunction of these effects.

The BH data indicate a self-instructional access ordering imposed by *S* upon his response set. The operation of self-instructional sets in the choice RT situation seems relatively well established. Hyman (1953) has noted the operation of what he called "verbal expectancies." More formally, one published article (Bernstein & Reese, 1965) and two as yet unpublished studies conducted by the present senior author, have all found facilitory effects, primarily occurring when U(r) exceeds 2 bits. Tentatively, it would appear that self-instruction sets can be understood in terms of the uncertainty reduction they afford. Uncertainty measures are at a maximum for a fixed number of alternatives when these alternatives are equiprobable. By adopting a self-instructional set, subjective-response probabilities are shifted away from the objective equiprobability and *S*s operate with levels of U(r) below that which is nominally imposed. This interpretation is congruent with Berlyne's (1957) position regarding relation between uncertainty and conflict. Hence, set would serve as a conflict minimizing operation that allows an *S* to operate more efficiently within response channel capacity constraints.

REFERENCES

Berlyne, D. C. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. *Psychol. Rev.,* 1957, **64,** 205–215.

Bernstein, I. H., & Reese, C. Behavioral hypotheses and choice reaction time. *Psychon. Sci.,* 1965, **3,** 259–260.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.,* 1962, **63,** 105–110.

Dillon, P. J. Stimulus versus response decisions as determinants of the relative frequency effect in disjunctive reaction time performance. *J. exp. Psychol.,* 1966, **71,** 321–330.

Dulany, D. E., Jr. The place of hypotheses and intentions; an analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness: A symposium of research.* Durham: Duke Univer. Press, 1962. Pp. 102–139.

Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1958.

Eriksen, C. W., & Collins, J. F. Reinterpretation of one form of backward and forward masking in visual perception. *J. exp. Psychol.,* 1965, **70,** 343–351.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.,* 1953, **45,** 188–196.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *J. exp. Psychol.,* 1964, **68,** 477–481.

Morin, R. E., Forrin, B., & Archer, W. Information processing behavior: the role of irrelevant stimulus information. *J. exp. Psychol.,* 1961, **61,** 89–96.

Posner, M. I. Information reduction in the analysis of sequential tasks. *Psychol. Rev.,* 1964, **71,** 491–504.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1955.

(Received August 2, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 525–537

EFFECT OF INITIAL NONREWARDED TRIALS: FACTORS RESPONSIBLE FOR INCREASED RESISTANCE TO EXTINCTION[1]

NORMAN E. SPEAR AND JOSEPH H. SPITZNER

Rutgers University

5 experiments investigated the source of the increased resistance to extinction of a running response occurring when rats receive no reinforcer subsequent to their initial responses. It was concluded that this source was similar to that contributing to the conventional partial reinforcement effect (PRE) but that it was not primarily attributable to habituation or to effects predicted by a proactive-interference model. Rather, the effect is a consequence of the lack of generalization decrement which otherwise facilitates rate of extinction.

If rats are given initial nonrewarded trials in a straight runway before receiving any rewards, their resistance to extinction is greater than that of *S*s given an equivalent number of trials, all rewarded (Spear, Hill, & O'Sullivan, 1965). However, this initial nonrewarded trials effect on extinction (INE) is small relative to the partial reinforcement effect on extinction (PRE) when number of rewards and nonrewards are held constant between the conditions (Spear et al., 1965). The INE and PRE, though differing in magnitude, represent similar effects that may be attributed to experience with nonrewarded trials regardless of the sequence of rewarded and nonrewarded trials preceding extinction. It is possible that the existence of the PRE and INE may be attributed to a common underlying process. If so, then their functional relationships with other variables should be equivalent. The first experiment of this paper dealt with a variable chosen for this purpose —retention interval. With a common underlying process, there was basis for expecting both a similar and a differential effect of this variable on the PRE and INE.

The PRE tends to decrease with an increase in retention interval (Aiken & Gibson, 1965). While the retention interval does act somewhat to decrease resistance to extinction after partial reinforcement (PRF), much of the effect apparently is due to an increase in resistance to extinction after continuous reinforcement (CRF) with the longer retention interval between acquisition and extinction. This latter fact has been independently demonstrated with a bar-press response (Youtz, 1938) and in human eyelid conditioning (Hilgard & Campbell, 1936). Aiken and Gibson have attributed these effects to a forgetting of the reinforcement contingencies of acquisition over the retention interval; their explanatory vehicle is the discrimination hypothesis of the PRE and the fact that the generalization gradient flattens as retention interval lengthens. Thus, *S*s given CRF increase their resistance to extinction after a longer retention interval from acquisition because the discrimination between acquisition and extinction has weakened. If this is an adequate explanation, then the INE and PRE should not be similarly affected by length of retention interval. For

---

[1] This research was supported in part by Grants MH-08888 and MH-12064 from the National Institute of Mental Health and a grant from the Society of Sigma Xi.

example, an *S* in a 24N–24R condition (24 initial nonrewarded trials followed by 24 rewarded trials) should have no more difficulty discriminating extinction from acquisition than an *S* given only 24 CRF trials (24R) prior to extinction. There would be no reason to predict that the greater resistance to extinction of Group 24N–24R over Group 24R (or 48R) should depend on the length of the interval between acquisition and extinction—that is if one function of this retention interval is to make *S* forget, and thereby neutralize, the effect of reinforcement schedule in acquisition on extinction. On the other hand, if the INE were reduced by increasing the retention interval, the explanation of Aiken and Gibson would appear inadequate. Moreover, this result would further imply that common processes underlie the INE and PRE.

EXPERIMENT I

The design was similar to that employed by Spear et al. (1965). The INE was defined in terms of the greater resistance to extinction of Group 24N–24R over two CRF groups —one equated for total number of acquisition trials (48R) and one for total number of rewarded acquisition trials (24R). To determine the effect of number of initial nonrewarded trials, a fourth group, Group 12N–24R, was also included. The PRE was defined in terms of the differential resistance to extinction of the CRF groups and Group 50%, given 24 rewarded trials interspersed with 24 nonrewarded trials in a quasirandom sequence. Retention interval was varied orthogonally to these five reinforcement conditions. Randomly selected *S*s from each reinforcement condition were given extinction either 1 or 8 days after the final acquisition session.

Method

Subjects and apparatus.—The *S*s included 10 randomly determined groups of 12 naive, female, albino rats, 60–80 days old, of the Sprague-Dawley strain. The apparatus was a gray, straight-alley runway with interior height and width of 3½ in., which provided the following running times (converted to speeds): Start—from the opening of the start-box door to a photobeam, 12 in. down the alley; Running—from the first beam to a second beam just before the goal-box entrance 22 in. away; Goal—from the entrance to the far corner of the goal box (14 in. wide and 6 in. high) where the reward container was kept, a distance of 12 in.; and Total—sum of start, running, and goal times.

Procedure.—All *S*s were handled and placed on the feeding schedule 7 days prior to the initial runway experience of the groups which later had an 8-day interval elapse between acquisition and extinction (Int groups). The initial handling and taming, as described by Spear et al. (1965), involved no reward pellets. About 20 min. after the end of each day's handling or runway session, *S* was fed 9 gm. of powdered lab chow plus a total of 18 45-mg. Noyes reward pellets (minus the number of pellets consumed in the runway).

Six trials were given per day. On a "rewarded" trial, 3 45-mg. Noyes pellets were placed in the glass food cup. On "nonrewarded" trials, the empty cup was present. During extinction, only nonrewarded trials were given; otherwise, treatment was identical to acquisition. The *S*s were run in rotation of 15 *S*s—3 from each acquisition group, including at least 1 from each of the 10 possible acquisition-extinction interval combinations. The minimum intertrial interval (ITI) for each *S* was held at 5 min. during acquisition and extinction. On each trial, the start-box door was opened 3 sec. after *S* had been placed in the start box facing away from the door. Upon entering the goal box, *S* was detained by closing the goal-box door for a minimum of 10 sec., or until all available pellets were consumed. If a clock was not stopped within 3 min., the time was recorded as 180 sec.; and *S* was placed next to the food cup.

Design.—The 2 × 5 factorial design included five major acquisition conditions and two intervals between acquisition and extinction. One set of five groups was given extinction trials 1 day (24 hr.) subsequent to the last acquisition session (Immediate

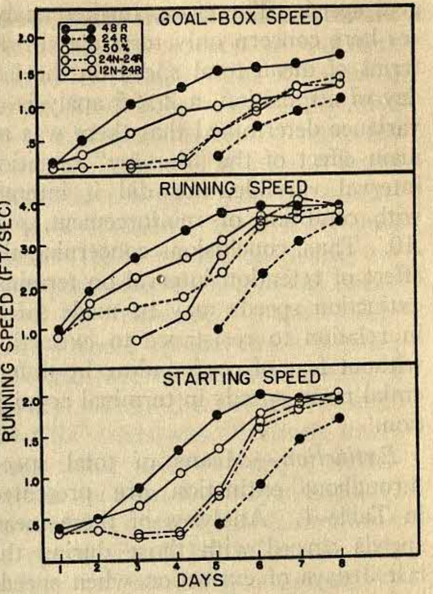


FIG. 1. Daily acquisition performance in Exp. I in terms of mean component speeds. (Since the retention variable had not yet been introduced, Int and Imm groups are combined in this figure.)

groups—Imm), and the other set of five groups began extinction 8 days after acquisition (Interval groups—Int). The acquisition treatments were as follows: Groups 48R received 48 rewarded trials, 6 on each of Days 1–8; Groups 50% received 24 rewarded interspersed with 24 nonrewarded trials, 6 on each of Days 1–8, according to the partial-reinforcement (PR) schedule described by Spear et al. (1965); Groups 24R received 24 rewarded trials, 6 on each of Days 5–8; Groups 12N–24R received 12 nonrewarded trials, 6 on each of Days 3 and 4, and 24 rewarded trials, 6 on each of Days 5–8; and Groups 24N–24R received 24 nonrewarded trials, 6 on each of Days 1–4, and 24 rewarded trials, 6 on each of Days 5–8. To equate for handling, *S*s in Groups 24R and 12N–24R were removed from their cages and held for 5 sec. during their turn in rotation on Acquisition Days 1–4 and 1–2, respectively. The Imm groups began acquisition the day after the Int groups completed acquisition trials, and all 10 groups were then given extinction (6 trials per day) on the same 6 days. The original feeding schedule was maintained for all *S*s during the 31 days elapsing from handling through extinction.

Thus, all groups were equated for amount of handling, degree of deprivation and body-weight depletion, and (except for Group 48R) number of rewarded trials prior to their common extinction experience.

Results and Discussion

Acquisition.—Three facts concerning the component speeds (see Fig. 1) are notable. First, differences in terminal acquisition (Day 8) as a function of condition of reinforcement were most apparent nearest the goal box. Condition of reinforcement during acquisition produced no significant differences in terms of mean start and running speed, but mean goal speed did differ among the Groups, F (4, 110) = 6.45, $p < .01$. The closer the goal, the greater the impairment to speed of running attributable to the nonrewarded trials experienced by Groups 50%, 12N–24R, and 24N–24R relative to Group 48R (see Fig. 1). On the other hand, the still slower speeds by Group 24R suggest that more benefit than detriment was attained as a consequence of these nonrewarded trials. Second, *S*s in Group 50% never ran faster during acquisition than those in Group 48R, although the tendency toward such a relationship was greatest with start speed and least with goal speed (cf. Spence, 1960, Chapter 6). Third, during initial nonrewarded trials there was a clear tendency for an increase in running speed, somewhat less of an increase in goal speed, and no increase in start speed. From the first to fourth day of initial nonrewarded trials, Group 24N–24R increased in mean running speed, t (23) = 5.51, $p < .001$, and goal speed, t (23) = 2.91, $p < .01$. A similar increase was found in Group 12N–24R. Neither group showed significant increases in start speeds during the initial nonrewarded trials.

Generally, though, the major results did not differ with start, running, or goal speed. Therefore, further analyses here concern only total speed. In terms of mean total speed on the last day of acquisition, a 2×5 analysis of variance determined that there was no main effect of the "dummy" retention interval variable, nor did it interact with condition of reinforcement, $p > .10$. Thus, conclusions concerning the effect of retention interval on terminal extinction speeds may be made safely in relation to resistance to extinction without fear of confounding by differential mean speeds in terminal acquisition.

Extinction.—Means of total speed throughout extinction are presented in Table 1. Analyses of these mean speeds agreed with those during the last 3 days of extinction when speeds began leveling off. The one exception is not critical to our conclusions (*S*s from Group 48R had greater mean speed throughout extinction than *S*s from Groups 24R, but there was no difference in terms of their mean speeds during the last 3 days of extinction). Also the results of these analyses of absolute speed were confirmed by analyses of mean trials-to-extinction criteria. It may be seen in Table 1 that Groups 48R and 24R, not previously experiencing nonrewarded trials, had greater speed in extinction after the longer retention interval between acquisition and extinction; the remaining groups, having nonrewarded trials during acquisition, showed relatively little change. Combining Groups 48R and 24R, *S*s extinguished after 8 days had reliably greater speeds than those extinguished after 1 day, $F\ (1, 44) = 12.33$, $p < .001$. Similar comparison in Groups 24N–24R and 12N–24R revealed no reliable differences in extinction speed as a function of interval between acquisition and extinction, $F\ (1, 44) < 1$. Group 50% actually had slightly greater speeds after 1 day than 8 days, although this difference was not statistically reliable. However, because of the increased extinction speeds by *S*s in Group 48R, it might be said that the conventional PRE did not occur when the 8-day interval was interpolated between acquisition and extinction. That is, neither in terms of mean speed throughout extinction nor during the last half of extinction did Group 50% differ at the .05 level (Duncan's multiple-range test) from Group 48R after the 8-day interval.

TABLE 1

Mean Total Speed (ft/sec) Per Trial during Extinction in Exp. I

	48R	24R	12N–24R	24N–24R	50%
Immediate					
M	.71	.54	.87	.96	1.37
SD	.20	.26	.25	.28	.50
Interval					
M	1.07	.78	.93	.96	1.34
SD	.45	.18	.37	.50	.41

The effect of retention interval on resistance to extinction was also examined by combining *S*s from Group 24N-24R with those from Group 12N-24R in comparison with combined *S*s from 24R and 48R. This also revealed overall greater resistance to extinction by *S*s given initial nonrewarded trials, $F\ (1, 92) = 5.69$, $p < .025$, and greater speed after the longer retention interval, $F\ (1, 92) = 5.13$, $p < .05$. However, the interaction, just missing statistical significance at the .05 level, $F\ (1, 92) = 3.72$, suggested that this latter effect was almost entirely due to *S*s not receiving initial nonrewarded trials. Thus, we conclude that the CRF Groups 48R and 24R increased their resistance to extinction after the longer retention interval between acquisition and extinction, but Groups given nonrewarded trials prior to extinction, whether interspersed among rewarded trials (Group 50%) or presented before any experience with rewarded trials at all (Groups 24N-24R and 12N-24R), did not. An important consequence was that the INE, present when extinction began 1 day after acquisition—mean speed of combined Groups 24N-24R and 12N-24R was greater than 24R and 48R, F (1, 44) = 15.87, $p < .001$—did not occur after the 8-day interval, F (1, 44) < 1.

These results are not consistent with predictions based upon the discrimination hypothesis of Aiken and Gibson (1965). Both the groups given initial nonrewarded trials had 24 successive CRF trials immediately preceding extinction. This would appear to put them equal to Group 24R and nearly equal to Group 48R in terms of their opportunity to discriminate acquisition from extinction. In spite of this, the groups given initial nonrewarded trials did not behave as did the CRF groups in responding to the increased interval between acquisition and extinction. Rather, as with Group 50%, their resistance to extinction was not increased by lengthening the interval between acquisition and extinction.

EXPERIMENT II

At this point, interest was diverted to the fact that groups given initial nonrewarded trials appeared to behave more like PRF *S*s than CRF *S*s, at least with respect to the effect of retention interval on resistance to extinction. Perhaps the influence of nonrewarded experiences pervaded the sequence of 24 successive CRF trials. The mechanism for this could have been proactive interference (PI) operating in Group 24N-24R in this way: subsequent to each rewarded experience, *S*'s representational response (RR) of the reward replaced *S*'s RR of nonreward; but during the interval between rewarded trials, the RR of nonreward spontaneously recovered to interfere with that for reward. The RR may be conceptualized simply as a cognitive unit—*S*'s presumably neural representation of the condition of reinforcement available in the goal box—or specifically as a fractional anticipatory goal response. It is viewed as having response properties so it may be extinguished (or "unlearned"), and it subsequently could spontaneously recover in strength over an interval of relative inactivity. Thus the successive series of 24 CRF trials was effectively interspersed with nonrewarded "experiences" in the form of *S*'s spontaneously recovered RR of the initial nonrewards. The operation of PI effects in this paradigm (a possibility suggested by Spear et al., 1965) was conceived to proceed in the following manner:

After initial nonrewarded trials, the RR of nonreward will be unlearned and replaced by the RR of the subsequent reward experience in the goal box. If, however, a sufficient temporal interval elapses between successive rewarded trials, the RR of the nonreward may be expected to recover and form part of the stimulus complex for the subsequent rewarded trial. Thus, subsequent to such rewarded trials, the RR of nonreward will acquire a certain capacity to evoke the instrumental response (perhaps it would be more fitting to say that the stimulus consequence of the RR of nonreward acquires this capacity). Subsequently when nonrewarded extinction trials are given, this acquired capacity will act specifically to maintain running and, in turn, greater resistance to extinction.[2]

Thus, this conception considered four

---

[2] This is little more than a slight modification of Capaldi's aftereffects theory (e.g., Capaldi, 1966), except that the present emphasis is on proactive, rather than retroactive, influences.

basic events with regard to the effects of initial nonrewarded trials. First, when a rewarded trial subsequently occurs, the RR of nonreward presumably will be unlearned and replaced by the RR of reward. Second, the unlearned RR of nonreward may be expected to spontaneously recover in strength during the intertrial interval (ITI) and thus interfere on a subsequent trial with the RR of reward. Third, we may expect *S* to respond in accord with the proportion of reward to nonreward RRs which form *S*'s RR of the condition of reinforcement present on that trial. Finally, as the stimulus consequence of the RR of nonreward becomes associated with the occurrence of reward, it will acquire a capacity to evoke the instrumental running response. Therefore, the greater the recovery during the ITI of the unlearned RR of nonreward, the greater will be the capacity of its stimulus consequence to evoke and maintain the instrumental running response during the nonrewarded trials of extinction.

The critical variable in this interpretation of the INE is the degree to which the RR of nonreward is permitted to recover between trials. One way to vary this is in terms of ITI. Before proceeding with conjectures of PI, it was desirable to directly test the consequences of such an interpretation. Therefore, Exp. II was designed to test the relative INE with two ITIs, 15 sec. and 15 min. It was predicted that the INE for groups given the shorter recovery time (ITI) would be slight relative to the groups permitted 15 min.

Method

Subjects and apparatus.—The *S*s were 64 (16 were randomly assigned to each group) experimentally naive female albino rats, 60–70 days old, of the Sprague-Dawley strain. The apparatus was a glass-enclosed, straight-alley runway (Hunter Manufacturing Company). Two running times (converted to speeds) were taken. Running speed was measured from 4½ in. beyond the start box to 7 in. before the goal-box entrance, a total of 21 in. Goal speed was measured from the end of running speed until *S* entered the goal box, a total of 7 in.

Procedure.—The procedure was identical to Exp. I with the following exceptions: (*a*) In addition to *S*'s daily ration of 9-gm. powdered chow, a total of 32 45-mg. Noyes pellets (minus those consumed in the runway) were given each *S* in its home cage about 20 min. after the day's runway experience. (*b*) The presence of *two* reward pellets defined a rewarded trial. (*c*) Sixteen trials were given to each *S* on each of 6 successive days. The *S*s given massed trials received 1 trial every 15 sec., and *S*s given spaced trials received 2 "massed" trials every 15 min. That is, the spaced-trial groups had only a 15-sec. wait in the holding cage between Trials 1 and 2, 3 and 4, 5 and 6, etc., of each day, but a 15-min. wait in the holding cage between Trials 2 and 3, 4 and 5, 6 and 7, etc., of each day. Thus, under spaced conditions, each odd-numbered trial followed a 15-min. interval, and each even-numbered trial followed a 15-sec. interval. One advantage of this method is that the effect of trial distribution may be observed within *S*s as well as between *S*s. Water was always available during the ITI, and a given *S* remained on a single ITI schedule throughout all stages of runway experience.

Design.—The two levels of ITI were given to independent groups treated as were Groups 48R and 24N-24R of Exp. I. The four groups of the 2 × 2 factorial design may be labeled in terms of treatment during acquisition (48R or 24N–24R) and ITI (massed—M or spaced—S trials). The groups, then, were: 48R-M, 48R-S, 24N–24R-M, and 24N–24R-S.

Results and Discussion

Acquisition.—Mean running speeds for each of the three stages of training may be seen in Fig. 2. The behavior of *S*s in Group 24N-24R-S during initial nonrewarded trials was notable. These *S*s ran nearly as fast as Group 48R-S *S*s on their odd trials (following a 15-min. interval), but nearly as slow as *S*s in Group 24N-24R-M on

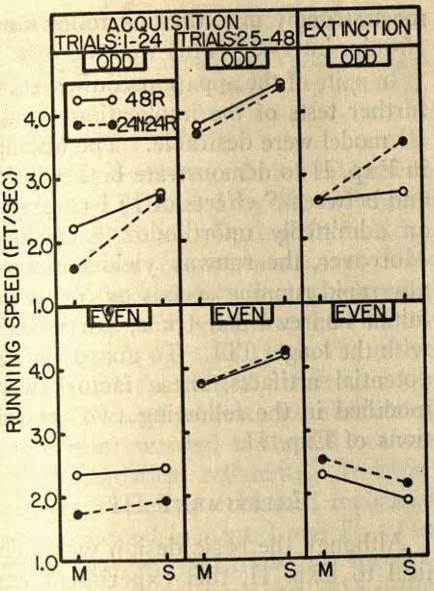


FIG. 2. Running speed for each group in each of the three stages of Exp. II. (Odd-numbered trials, preceded by a 15-min. ITI for spaced *S*s, and even-numbered trials, preceded by a 15-sec. ITI for both massed and spaced *S*s, are shown separately.)

their even trials (following a 15-sec. interval). This was not the case in terms of goal speed where *S*s in Groups 24N-24R ran consistently slower on all of their initial nonrewarded trials than *S*s in either of Groups 48R, regardless of the ITI. In all other respects, however, the results in terms of goal speeds agreed with those of running speeds, so only the latter (less variable) measure will be considered in detail.

Separate analyses of variance were performed on the even and odd trials. During odd Trials 1–23 before reward was introduced in the 24N-24R groups, speeds were greater for the spaced-trials groups, F (1, 60) = 61.90, $p < .001$; and presence vs. absence of reward made a greater difference in mean speed under massed-trial conditions than with spaced trials. This latter result was reflected by the significant interaction between distribution of trials and reward conditions, F (1, 60) = 7.23, $p < .01$, together with the overall superiority in the running speed of Groups 48R, F (1, 60) = 11.52, $p < .001$. On even Trials 2–24, however, spaced-trial *S*s ran no faster than massed-trial *S*s, F (1, 60) = 2.14, $p > .05$. Since *all* *S*s received their even trials 15 sec. after their preceding trial, this suggests that the effect of distribution of trials is quite temporary.

When reward was introduced to Groups 24N-24R (Acquisition Trials 25–48) the mean running speeds of these groups soon did not differ from those of Group 48R. That is, presence vs. absence of reward on Trials 1–24 did not result in a significant difference in running speed on either odd or even Trials 25–48, nor did this variable interact significantly with trial distribution during these trials (*F*s ranged from .10 to .84). However, *S*s given spaced trials continued to run markedly faster than those given massed trials on both odd trials, F (1, 60) = 35.70, $p < .001$, and even trials, F (1, 60) = 16.71, $p < .001$.

Extinction.—As in acquisition, the slightly more variable goal speed yielded the same relationships as did running speed, so only the latter will be considered in detail. It is apparent in Fig. 2 that the predicted interaction was thoroughly confirmed on the odd trials. An analysis of variance of mean running speed on odd Trials 1–47 in extinction substantiated the reliability of this interaction, F (1, 60) = 8.73, $p < .005$, produced by the occurrence of considerably faster speeds by Group 24N-24R-S than Group 48R-S, in contrast to the negligible differences between the mean speed of Groups 24N-24R-M and 48R-M. In other words, the INE was reliably greater with spaced than with massed trials. Both the main effects of trial distribution, F (1, 60) = 16.76, $p < .001$, and of the differential reward conditions in acquisition, F (1, 60) = 9.59, $p < .005$, were also statistically significant over these trials.

The same analysis completed on even Trials 2–48 revealed quite different relationships. No trace of interaction was found, F (1, 60) < 1; the greater speed by the 24N-24R groups compared to 48R did not quite attain statistical significance at the .05 level, F (1, 60 = 3.34, $p < .10$; and *S*s given massed trials on all trials had numerically, though unreliably, *greater* speeds than those given spaced trials between the massed trials under analysis here.

The results of Exp. II thus agreed with the predictions of the PI model in terms of performance on the odd extinction trials: the INE was present more strongly the longer the ITI. The fact that this effect was not evident on even trials is also expected from the model. Recall that the INE is primarily attributed to the spontaneous recovery of the previously "unlearned" RR of nonreward during the ITI and the consequent reinforced pairing of its stimulus consequence with instrumental running behavior. Recovery and subsequent conditioning of the nonreward RR must be most pronounced following the longer ITI, i.e., on an odd trial. It is to be expected, then, that this conditioning should be most apparent the greater the commonality of stimulus elements from "conditioning" (rewarded acquisition trials) to "test" (extinction) trials. That is, to the extent that the stimulus consequence of nonreward RR was conditioned to running primarily on trials following a 15-min. interval, this conditioning should have been reflected most strongly in extinction following a 15-min. interval. This was the case.

In spite of the apparent confirmation, further tests of the implications of the PI model were desirable. The attempt in Exp. II to demonstrate both within- and between-*S* effects of ITI employed an admittedly unorthodox procedure. Moreover, the runway yielded unusually rapid running speeds by *S*s on the initial nonrewarded trials, particularly with the longer ITI. To guard against potential artifacts, these factors were modified in the following two replications of Exp. II.

EXPERIMENT III

Although the basic design was identical to Exp. II, this experiment employed a different runway, an additional speed measure (starting speed), and a more conventional means of varying ITI.

Method

Subjects and apparatus.—The *S*s were 40 naive female albino rats of the Sprague-Dawley strain, approximately 65 days old, weighing 180–200 gm. The straight runway and response measures—start, running, and goal times—were employed as reported by Vogel, Mikulka, and Spear (1966). An added measure, total time, was the combined start, running, and goal times. Times were converted to reciprocals and reported as speeds in ft/sec.

Design.—The four experimental groups of 10 randomly assigned *S*s constituted a 2 × 2 factorial design varying reward during the first 24 of the 48 acquisition trials and ITI. Half of the *S*s were returned to home cages for 15 sec. between trials (massed—M); the other half for 20 min. (spaced—S). Six trials were given each day. The same ITI was given a single *S* throughout training. For half the *S*s run under each ITI, the first 24 acquisition trials were nonrewarded. Acquisition Trials 25–48 (24N–24R) used as reward 2 45-mg. Noyes pellets. The second half of *S*s run under each ITI were rewarded with 2 pellets on each acquisition trial (48R). The four groups were: 48R-M, 48R-S, 24N–24R-M, and 24N–24R-S.

All groups received 30 extinction trials immediately following acquisition.

Otherwise, the experimental procedure was essentially identical to that of Exp. I and II.

Results and Discussion

Figure 3 shows that the critical interaction predicted by the PI model was not obtained. During the last day of extinction (representative of relationships throughout) total speeds were greater for *S*s given initial nonrewarded trials, $p < .001$, but the effect of ITI and the critical interaction between ITI and presence vs. absence of initial nonrewarded trials did not approach statistical reliability. In other respects, the effects obtained replicated those obtained in Exp. I and II. The initial nonrewarded trials had increasingly inhibitory effects on subsequent acquisition the closer *S* came to the goal box: on the last day of acquisition, Groups 24N-24R had reliably greater running speeds than Groups 48R, $p < .05$, and nearly reliably greater start speeds, $p < .10$; but the absence of differences in goal speed resulted in total-speed differences not statistically significant, $p > .10$. On the last day of extinction, as in Exp. I and II, the greater speeds by Group 24N-24R were most apparent in the goal and running speeds. However, at no point was the effect of ITI as great as it had been in Exp. II. It appears that *S*s given alternate experience of running under massed- and spaced-trial conditions reflect an exaggerated effect of trial spacing. Similarly, exaggerated effects of reward magnitude have been reported when *S*s are given quasi-alternate experience with small and large reinforcers. Thus, variation in reward magnitude has a greater effect on acquisition behavior if the variation is within *S*s—that is, if a given *S*'s performance is measured as *S* is exposed to two or more different reward magnitudes—than between *S*s (e.g., Schrier, 1958; Spear & Spitzner, 1966). Apparently, this is also true for the effect of ITI.

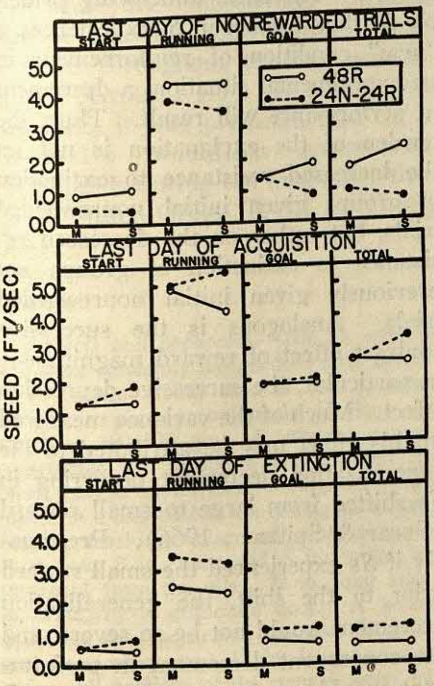


FIG. 3. Component and total speeds during the last day of each experimental stage for each group in Exp. III.

Therefore, Exp. III failed to confirm the results obtained on the critical trials in Exp. II and the implication of the PI model that the INE should be greater with spaced than with massed trials. An additional experiment, designed to destroy the balance between Exp. II and III, differed in three respects: (*a*) A different runway was used; (*b*) Three levels of ITI were included (15 sec., 2 min., and 30 min.) forming a 2 × 3 factorial design; and (*c*) The ITI variable was introduced only during Trials 25–48 of acquisition where, according to the PI model, the longer ITI should permit greater operation of processes which increase subsequent resistance to extinction.

The results of this experiment agreed with those of Exp. III: Groups 24N-24R were more resistant to extinction than Groups 48R (though the absolute effect was smaller than previously), but this effect was not increased with increasing ITI. Thus, the results of this experiment were not in accord with the PI model.

Of course, the fact that the prediction of the PI model was not upheld in this situation does not imply that it is generally inappropriate. Since all trials were not given within a single session in these experiments, and since 24 hr. elapsed between sessions, opportunity for recovery was present between sessions as well as between trials. Perhaps sufficient recovery occurred between sessions to effectively mask the effect of the conventional variation in ITI given in Exp. III and the experiment described above. Furthermore, other research has suggested that the characteristics of PI may be markedly different if several learning sessions are each separated by 24 hr. than if all learning occurs in a single day's session (cf. Kehoe, 1963; Underwood & Ekstrand, 1966).

Regardless, it had become evident that other explanations of the INE might be more useful within the procedures employed. There appeared to be two major candidates. First, the initial nonrewarded trials may have the effect of causing *S* to attend to, and thus *habituate* to, various stimuli in the runway situation. The *S*s rewarded on these trials might be presumed to focus attention on the goal box and traverse the runway faster, thus allowing less occasion to habituate to these competing stimuli. On subsequent extinction trials, the competing stimuli to which *S* had not previously habituated could elicit competing responses in *S*. Therefore, the groups without the initial nonrewarded trials would have more competing stimuli, competing responses, and consequently slower running speeds in extinction. On the basis of previous data (Spear, Hill, & O'Sullivan, 1963), this hypothesis was discarded. In this experiment, *S*s were given their initial nonrewarded experiences in a different apparatus and different experimental room than was employed for extinction. Thus it was unlikely that differential habituation of specific competing stimuli and competing responses were responsible for the INE.

A second explanation of the INE, simpler but perhaps more likely, is that initially nonrewarded *S*s run faster during extinction trials simply because they *previously* had experienced the condition of reinforcement present during the extinction trials—zero nominal reward. The basic underlying principle is that whenever *S* experiences a "new" condition of reinforcement in an experimental situation, a decrement in performance will result. Thus, the burden of the explanation is not on the increased resistance to extinction of groups given initial nonrewarded trials, but rather on the decreased resistance to extinction of groups *not* previously given initial nonrewarded trials. Analogous is the successive contrast effect of reward magnitude—in particular, the successive depression effect. Much of the variance measured in this effect may be attributed to the generalization decrement occurring in *S*s shifted from large to small reward (Spear & Spitzner, 1966). Presumably if *S*s experienced the small reward prior to the shift, the generalization decrement would not be so severe, and the consequent decrement in performance would not be so great. Experiment IV was designed to extend this principle to the INE.

EXPERIMENT IV

The intent was to replicate the basic paradigm of the INE; but instead of having the experimental group shifted nonrewards-rewards-nonrewards, the sequence was small reward-large-small. The reward values were chosen to be comparable in absolute and relative value to the zero and 2 45-mg. pellets employed in the INE paradigm (Exp. II and III above; Spear et al., 1965). In Exp. IV, small reward was 1 20-mg. pellet; large reward was 6 20-mg. pellets. Thus, Group 1-6-1 (initial 24 trials on 1 pellet, next 24 trials on 6 pellets, final trials on 1 pellet) corresponded to Group 24N-24R of the previous experiments. Also, Group H-6-1 (handled for initial 24 trials, given 6 pellets for the next 24 trials, and 1 pellet in the final trials) and Group 6-6-1 corresponded to Groups 24R and 48R, respectively. In addition, a base-line control group given 1 pellet throughout training was included (Group 1-1-1). In accord with the INE, it was predicted that *S*s in Group 1-6-1 would have greater running speeds during the final stage than *S*s in either Group H-6-1 or Group 6-6-1.

Method

Subjects and apparatus.—The *S*s were 40 (10 were randomly assigned to each group) naive female albino rats of the Sprague-Dawley strain, approximately 65 days old, weighing 180–200 gm. A straight runway constructed of Masonite and painted flat black, except for the clear Plexiglas top, was placed in a dim cubicle containing an exhaust fan for masking noise and ventilation. A start box, 8 in. long and 5 in. wide, was separated from the alley, 72 in. long and 3 in. wide, by a guillotine door. A second guillotine door separated the alley from the goal box, which was 12 in. long and 5 in. wide. All portions of the runway were 5 in. high. For purposes irrelevant to the present study, hardware cloth covered a 4-ft. middle section of the floor.

Response measures included start time (from the opening of the start-box door to a photobeam 6½ in. down the alley), running time (from the first beam to a second, 60½ in. farther and 6 in. before the goal-box door), and goal time (from the second beam to a third, 8 in. inside the goal box and 1 in. before the reward-pellet dish at the rear). Times were converted to reciprocals and reported as speeds in ft/sec, including total speed.

Design.—The *S*s were run in two replications of 20 *S*s each, each replication containing five members from each of the four groups. Analysis of the data included replications as a source of variance; but since this variable failed to interact with any other variable, it will not be mentioned further.

Training was divided into three continuous stages—Preshift I (4 days), Preshift II (4 days), and Postshift (5 days). The four randomly assigned groups were: 1-1-1, 1-6-1, 6-6-1, and H-6-1, with symbols representing the reward (number of 20-mg. pellets) received in each successive stage. Group H-6-1 received no runway training during Preshift I; each *S* was lifted from the cage and immediately replaced during its turn in the trial rotation. The ITI was about 5 min. throughout all stages of training.

Otherwise, except for a slight change in daily ration (10-gm. chow plus balance of 36 20-mg. pellets after consumption in the goal box), the procedure was essentially identical to the earlier experiments on the INE.

Results and Discussion

Unless noted otherwise, it may be assumed that the results were consistent for all speed measures; and analyses are presented only in terms of total speed. The conventional effect of greater speed for greater reward magnitude was obtained during Preshift I, $F\ (1,\ 24) = 9.74,\ p < .005$. Subsequent to the shift from 1 to 6 pellets in Preshift II, the total speed of Group 1-6-1 adjusted to the level of Group 6-6-1. As consistently found in our laboratory with this "elation" effect paradigm and with the INE paradigm (see, e.g., Exp. III of this report), there was some tendency for start and running speeds of Group 1-6-1 to overshoot the base-line level of Group 6-6-1. In fact, the overshooting in terms of running speed was reliable on Day 4, $F\ (1, 16) = 7.45$, $p < .025$, but no such tendency existed in terms of goal speed (see Fig. 4).

The results in Postshift completely conformed to the expectations of the INE. As may be seen in Fig. 4, the speeds in Group 1-6-1 showed little, if any, of the decrement occurring in 6-6-1 and to a lesser extent in H-6-1. In accord with the INE, the mean speed of 1-6-1 was greater than that of either 6-6-1 or H-6-1 during Postshift. This was the case whether comparison was made in terms of total speed throughout Postshift, $F\ (1, 16) = 5.00$ and 5.00, respectively, $p < .05$; or in terms of the last 2 days of this stage, $F\ (1, 16) = 10.23$ and 4.46, respectively, $p < .05$. The mean speed of Groups H-6-1 and 6-6-1 did not differ, nor did 1-6-1 and 1-1-1 (when these comparisons were made throughout all of Postshift and for the last 2 days, *F*s ranged from .01 to 1.18). In accord with the depression effect, the former two groups had slower speeds than the latter throughout Postshift, $F\ (1, 32) = 8.03$, $p < .01$, and during the last 2 days, $F\ (1, 32) = 12.94$, $p < .001$. Of course the analogy with the INE implies that Group 1-6-1 would not show a similar depression effect, and they did not: as noted above, their Postshift speeds did not differ reliably from those of Group 1-1-1. The absence of a depression effect in this group was also predicted from the results of Spear and Spitzner (1966, Exp. IV).

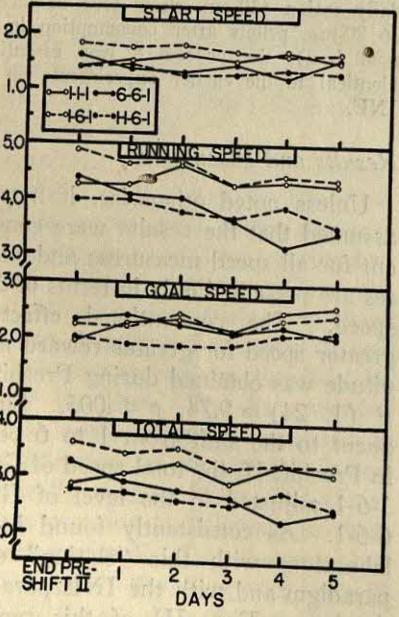


FIG. 4. Mean component and total speeds during the last day of Preshift II and each day of Postshift for each group in Exp. IV.

The greater speed during Postshift by Group 1-6-1 compared with 6-6-1 or H-6-1, together with the INE, elicits the following generalization: speed of running for a particular condition of reinforcement is greater for *S*s previously given it than for *S*s without this prior experience. This is most likely a consequence of a decrement occurring in the performance of *S*s faced with a new condition of reinforcement, perhaps akin to the decrement found in performance whenever any new stimulus is introduced. This generalization accommodates the phenomena in the present experiments fairly well and is consistent with the effects of shifts in reward magnitude in other situations. It is particularly adept at handling one phenomenon reported by Spear and Spitzer (1966; Exp. IV). It was shown that *S*s shifted from large to small reward associated with one stimulus were less likely to show a depression effect if they had previously experienced that small reward—even though prior experience with the small reward was in association with a quite different stimulus. While the PI model may be employed to advantage in similar situations, it is clearly not as useful for explaining this phenomenon and those reported in the present paper as is the generalization emerging from Exp. IV.

REFERENCES

AIKEN, E. G., & GIBSON, K. L. Continuous and fixed ratio reinforcement effects in extinction one day and three weeks after acquisition. *Psychon. Sci.*, 1965, **3**, 527–528.

CAPALDI, E. J. Partial reinforcement: A hypothesis of sequential effects. *Psychol. Rev.*, 1966, **73**, 459–477.

HILGARD, E. R., & CAMPBELL, A. A. The course of acquisition and retention of conditioned eye-lid responses in man. *J. exp. Psychol.*, 1936, **19**, 227–247.

KEHOE, J. Effects of prior and interpolated learning on retention in pigeons. *J. exp. Psychol.*, 1963, **65**, 537–545.

SCHRIER, A. M. Comparison of two methods of investigating the effect of amount of reward on performance. *J. comp. physiol. Psychol.*, 1958, **51**, 725–731.

SPEAR, N. E., HILL, W. F., & O'SULLIVAN, D. J. Extinction after "frustrating" and "nonfrustrating" nonreward. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, 1963.

SPEAR, N. E., HILL, W. F., & O'SULLIVAN, D. J. Acquisition and extinction after initial trials without reward. *J. exp. Psychol.*, 1965, **69**, 25–29.

SPEAR, N. E., & SPITZNER, J. H. Simultaneous and successive contrast effects of reward magnitude in selective learning. *Psychol. Monogr.*, 1966, **80**(10, Whole No. 618).

SPENCE, K. W. *Behavior theory and learning: Selected papers.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

UNDERWOOD, B. J., & EKSTRAND, B. R. An analysis of some shortcomings in the interference theory of forgetting. *Psychol. Rev.*, 1966, **73**, 540–549.

VOGEL, J. R., MIKULKA, P., & SPEAR, N. E. Effect of interpolated extinction and level of training on the "depression effect." *J. exp. Psychol.*, 1966, **72**, 51–60.

YOUTZ, R. E. P. The change with time of a Thordikian response in the rat. *J. exp. Psychol.*, 1938, **23**, 128–140.

(Received August 3, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 538–542

INFLUENCE OF INSTRUCTIONS ON DEGREE OF SHAPE CONSTANCY[1]

WILLIAM H. LICHTE AND C. ROBERT BORRESEN

University of Missouri *Wichita State University*

The Ss were instructed to respond either to "real" shape (object instructions), how the stimulus looked (apparent instructions), or the stimulus silhouette (image instructions). 3 groups of 15 Ss were used, each with only one kind of instruction. The 4 standard stimuli (SS) were complex and unfamiliar. The S manipulated the shape of the variable stimulus (VS) to equal that of the turned SS. Obtained Brunswik ratios (BRs) for the Object, Apparent, and Image groups were .90, .62, and .45; the first differed significantly from the others. Distributions of individuals' mean BRs for Object and Image groups did not overlap; that of the Apparent group was U shaped and overlapped the others. Results suggested that, in a "normal" situation, (a) object instructions produce nearly "perfect" constancy, (b) the image instruction cannot eliminate constancy, and (c) the conventional apparent instruction is ambiguous, resulting in either an object or image set.

In the performance of the shape-constancy task in an experiment, there are two easily identified kinds of responses *S* may show. One is to indicate the shape of the stimulus object when optically projected upon a plane perpendicular to *S*s' line of sight (the retinal *image* value), the second is the response indicating the shape of the stimulus object when it is placed perpendicular to the line of sight (the *object* value). There is a third, less clearly defined, shape-constancy task which is typically described as giving the "perceived" or "apparent" shape of the object. It is difficult to define this further, other than to add that responding to the immediate impression is usually stressed.

In his early work in shape constancy Thouless (1930–31) explicitly recognized the first two kinds of responses as opposite ends of a continuum. His instruction to *S*s indicated his assumption that the shapes experienced would fall somewhere between these poles. His data bore out this assumption.

Studies on the effects of instructions have not been numerous. Klimpfinger (1933) obtained a large number of judgments from five highly sophisticated *S*s, each of which responded under *synthetic* and *analytic* conditions. In the first, *S*s were to respond to stimuli as *objects,* in the second as *pictures* or *projections*. The sophistication of *S*s and the use of both instructions with each may be faults in the study. Under the *synthetic* condition constancy was high—mean Brunswik ratio (BR) was .73—and in the *analytic* condition it was low—mean BR was .28.

Instructions were one of the experimental variables in a study by Epstein, Bontrager, and Park (1962). The three kinds of instructions were labeled *phenomenal, objective,* and *analytic.* The task in the latter was clearly explained as one calling for a reproduction of the retinal-image shape; the objective instruction called for a reproduction of the "actual physical shape" of the stimulus object. The phenomenal instruction was "to adjust the shape . . . so that it appears to be the same as . . ." The immediate perceptual impression was also stressed. Computation from the published data for one condition (vertical background) gives BRs of .21, .30, and .58 for analytic, phenomenal, and objective instructions, respectively.

---

[1] This study was reported in a paper read at a meeting of the Midwestern Psychological Association in St. Louis in 1960. It was supported in part through a summer research appointment given the first author by the University Research Council, University of Missouri, and in part by Contract AF33-(600)31315, held by IBM Corporation, University of Missouri subcontract.

These studies indicate that an instruction to respond to the stimuli as objects gives the greatest degree of (though far from perfect) shape constancy. A factor which may be of even greater importance than the formal instructions in influencing *S*'s responses is the implication of the total experimental situation and the procedure. This will be discussed further on.

The general purpose of the present study was to determine the effects of the three kinds of instructions described at the beginning of this paper. They will be referred to as image, apparent, and object instructions. More specifically, the experimental hypotheses were that object instructions would give the highest degree of constancy, the image instructions the lowest degree, and the apparent instructions an intermediate degree; and that the variability of responses with the apparent instructions would be greater than for either of the other two.

The conditions of the experiment reflect certain attitudes on the part of the writers. One of these was that the shape-constancy phenomenon relates to realistic, every day type of perceiving and that, insofar as possible, the situation and the looking behavior of *S* should be of this kind. Another was that the use of simple geometrical figures is an unnecessary and undesirable limitation, likely to make the task seem so simple to *S* that he would tend to justify it through over-elaborate interpretations. For this reason, the stimulus figures were complex: many-sided, unsymmetrical, and relatively meaningless.

METHOD

Apparatus.—The apparatus and physical situation were identical with those in a previous study by Borresen and Lichte (1962). The *S* sat facing three adjacent, vertical panels, each about 1.2 m. from his forehead. Each panel had, at eye level, a circular aperture 18 cm. in diameter. The Standard-Stimulus (SS) object, a white paper on a stiff backing, appeared in the center aperture. The *E* could adjust the SS object to any degree of rotation on a vertical axis through its center. In the left aperture was a disk of approximately the same area as the standard stimuli; *S* could change the angle of rotation of this disk by turning a crank near his left hand. In the right aperture was the Variable Stimulus (VS); this was a patch of white light projected from the rear on a translucent paper screen. The *S* could change the shape of the VS (expanding or contracting it horizontally) by turning a crank near his right hand.

From the same previous study, Stimuli A, B, C, and D were used. Stimulus E, which gave a high degree of constancy in the first study, was omitted to influence the results against the expectation of high constancy with a clearcut object instruction.

Subjects.—The *S*s were 45 university summer-school students, male and female. The median age of *S*s was 21, the range was 17–36.

Design.—There were three experimental groups of 15 *S*s each. The *S*s were placed in the three groups in turn, as they appeared for duty. The *S*s in each group received one, and only one, type of instruction.

Procedure.—The *S*s were run individually, each in a single session which lasted about 50 min. In the first part of the session *S* was given a small amount of familiarity with the "nonsense" stimuli. The purpose was to reduce the within-*S* variability which would result if constancy grew rapidly in the first few trials, as the data of Borresen and Lichte (1962) indicate it would. Familiarity was given by having *S* make judgments about the slant of the standard stimuli. Each of the four stimuli was presented seven times, once each at slants (with reference to the line-of-sight) of 40°, 45°, 50°, 55°, 60°, 65°, and 70°, and *S* rotated the circular disk in the aperture on his left until its slant appeared equal to that of the standard stimulus.

After these trials, *S* was told that the next part of the experiment would deal with the shapes of the stimuli. He was then read the instructions appropriate to the group to which he was being assigned. The three sets of instructions follow.

> *Image* instructions: When an object in the center stage is turned at some angle the perspective view (or the *image* of the object) has a different shape than when the object faces you directly. By the image or the perspective view we mean the light pattern that actually falls on the back of the eye or that would strike the film of a camera put in place of your eye. The image or perspective view is also equivalent to the shadow that would fall on the background if a light were substituted for your eye. Your task in this part of the experiment is to adjust the shape of the changeable form in the right stage until it is as much like the perspective view or image shape of the object in the center stage as possible. You may adjust the changeable form back and forth as much as may be necessary. Again, please use normal and ordinary methods of looking. Now, to be sure that we understand each other, will you please tell me in your own words what the task is.
>
> *Apparent* instructions: When the object in the center stage is turned at some angle the perspective view or the image shape of the object is different than when it is facing you directly. (The perspective view or image shape of the object is equivalent to the shadow that would be cast on the background if a light were substituted for your eye.) However, even though the object is turned, it remains an object with a real shape—the shape seen when it is facing you directly. Now when you look at the turned object the actual shape you experience might conceivably be that of the image, or that of the real object, or some compromise between the two. Your task in this part of the experiment is to reproduce, as nearly as possible, the shape you experience when you look at the object in the center stage. You reproduce this in the stage on the right by manipulating the changeable form there. Adjust the changeable form so that its shape is as much as possible like the shape you experience when you look at the object in the center stage. You may adjust . . .
>
> *Object* instructions: When the forms in the center stage are placed at various angles they give you, in a sense, different pictures or images. For example, if a light were substituted for your eye, the forms would cast differently shaped shadows on the background as they were placed in different positions. Despite this fact, when any form is placed in various positions it remains the same object with the same real shape. By real shape we will mean the shape that you see when the object is facing directly toward you. In this part of the experiment you will make judgments about this real shape. Your task will be to manipulate the shape of the changeable form in the right stage until you have duplicated as nearly as possible the real shape of the form in the center stage. You may adjust . . .

When *S* had stated the task in his own words, *E* made a judgment of his understanding and either accepted him as an *S* or tried to improve his understanding by rereading appropriate parts (or the whole) of the instructions. If these were still not understood, *E* decided at this point not to use the data. (Rejected *S*s, however, went through the whole experimental procedure.) The data of four *S*s were rejected; one of these *S*s had been given object instructions, one image instructions, and two apparent instructions.

Following the instructions, each *S* had 36 trials in which the shape of the VS was adjusted to match the SS. Each of the four standard stimuli was presented at three slants, 45°, 55°, and 65° from the line of sight, and each stimulus-slant combination was matched three times. Order of stimuli and order of slant were randomized.

RESULTS

A BR was computed for the mean of the three settings at each stimulus-slant combination, thus there were 12 BRs for each *S*. These data were treated by analysis of variance, and the results are shown in Table 1.

The between-groups variance was significant at less than the .01 level. The mean BRs for the Object, Apparent, and Image groups were .90, .62, and .45, respectively. The instructions variable significantly influenced the results and the rank order of the obtained BRs was in the predicted direction. The mean BR for the Object group differed significantly (< .01) from that of the Apparent group and the Image group. The latter groups did not differ significantly from each other. The mean BR of the Image group differed significantly (< .01) from zero.

The only within-groups variable significant at the .01 level was slant. The differences between means, however, were slight. The mean BRs for slants of 45°, 55°, and 65° were .68, .67, and .61, respectively.

A mean BR was computed for each *S*. The distributions of these means in the three groups are shown in Fig. 1. The curves for the Image and Object groups do not overlap, but the curve for the Apparent group is U shaped and overlaps each of the others. The variances for the Image, Apparent, and Object groups were .16, .36, and .16, respectively, but differences were not statistically significant.

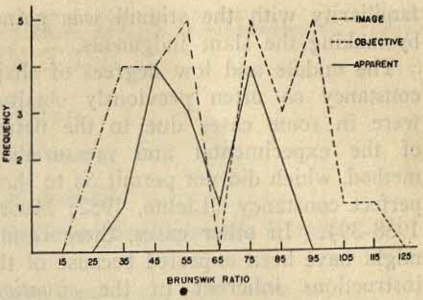


FIG. 1. Distributions of Brunswik ratios for the mean settings of the individual *S*s in the three groups.

TABLE 1

ANALYSIS OF VARIANCE OF BRUNSWIK RATIOS

Source	*df*	*MS*	*F*
Between *S*s	44		
Groups	2	9.52	50.11**
Error (b)	42	.19	
Within *S*s	495		
Figure	3	.05	1.67
Slant	2	.24	6.00**
F × S	6	.02	.67
G × F	6	.02	.67
G × S	4	.10	2.50*
G × F × S	12	.02	.67
Error (w)	462	.03	
Error_1	126	.03	
Error_2	84	.04	
Error_3	252	.03	

* $p = .05$.
** $p = .01$.

DISCUSSION

The results show that the instructions to *S* do make an important difference in his responses and they indicate particularly that an *object* instruction produces a high degree of constancy as compared with either of the other two types. This raises the question as to what level of constancy should be expected in realistic perceiving situations. Most studies have obtained middle or low degrees of shape constancy (BRs below .75). These results are in contrast to those in size constancy, which is usually nearly perfect. It is argued here that the high degrees of size constancy usually obtained are the result of the clear objective instructions commonly used.

The writers believe that a clear, thorough going, object instruction fits in with the meaning of shape constancy as usually defined in textbook and theoretical discussions, and that with such instruction, obtained degrees of shape constancy will usually be high in realistic or "enriched" situations. In the present study, high shape constancy was expected (in part) because of the realistic nature of the situation: head position was not rigidly fixed, binocular vision was used, there was weak general illumination, and familiarity with the stimuli was gained by making the slant judgments.

The middle and low degrees of shape constancy so often previously obtained were in some cases due to the nature of the experimental and measurement method, which did not permit *S*s to show perfect constancy (Lichte, 1952; Moore, 1938–39). In other cases, these results might have been expected because of the instructions inherent in the *situation*. The social-psychological influences in the experimental situation may have caused *S*s to "say things to themselves" (Farber, 1963) which effectively changed the instructions. When the august European professor put a disk on a table and asked *S*, "What does the shape appear to be?" it was probably immediately obvious to *S* that it was a circle and it must have seemed equally obvious that the professor would not go to all this trouble if the answer was as simple as that. The *S* then jumped to the conclusion that his task was much more elaborate, and as a result he probably operated in terms of a somewhat hazy concept of the retinal image. This factor may also have influenced the results in studies where the instructions were meant to be objective (as that term is defined here). Even instructions in the early studies intended to be of this type were not as thoroughly object centered as in the present study or in that of Epstein, Bontrager, and Park (1962).

This interpretation of the results of older studies and the data of the present experiment support the view that the apparent type of instruction should be avoided. The results with such an instruction are neither clear nor unique, and the data suggest that *S*s so instructed will operate either like those with an object or an image set.

It is possible that whenever simple and familiar geometrical figures (square, circle, diamond, "perfect" rectangle, equilateral triangle) are used as stimuli the shape may be so obvious that *S* cannot believe that he is to reproduce it and slips into some other personally defined type of set, as illustrated by Farber (1963) in another performance area.

The obtained BR of .45 with the image instructions means that with an enriched situation and moderately complex figures *S* is unable to respond only in terms of his retinal images. It seems that those conditions which produce shape constancy will produce it to some degree even with the most "analytical" or stimulus-oriented set.

The writers conclude that in shape-constancy experiments the instructions should be thoughtfully selected and should be either the image or object type, depending on the purpose of the study. The clarity and forcefulness of the instructions should be verified with *S*s. When more or less "normal" perception is the object of study, the instructions should be of the object type, and a high degree of constancy is to be expected.

REFERENCES

BORRESEN, C. R., & LICHTE, W. H. Shape-constancy: Dependence upon stimulus familiarity. *J. exp. Psychol.*, 1962, **63**, 91–97.

EPSTEIN, W., BONTRAGER, H., & PARK, J. The induction of nonveridical slant and the perception of shape. *J. exp. Psychol.*, 1962, **63**, 472–479.

FARBER, I. E. The things people say to themselves. *Amer. J. Psychol.*, 1963, **18**, 185–197.

KLIMPFINGER, S. Über den Einfluss von internationaler Einstellung und Übung auf die Gestaltkonstanz. *Arch ges. Psychol.*, 1933, **88**, 551–598.

LICHTE, W. H. Shape constancy: Dependence upon angle of rotation; individual differences. *J. exp. Psychol.* 1952, **43**, 49–57.

MOORE, W. E. Experiments on the constancy of shape. *Brit. J. Psychol.*, 1938–39, **29**, 104–116.

THOULESS, R. H. Phenomenal regression to the real object. I. *Brit. J. Psychol.*, 1930–31, **21**, 339–359.

(Received August 3, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 543–549

PERCEPTUAL SELECTIVITY IN A VISUAL RECOGNITION TASK[1]

HOWARD EGETH
Johns Hopkins University

AND EDWARD E. SMITH
Behavioral and Clinical Studies Research Center, St. Elizabeth's Hospital

The present experiment is an attempt to localize the mechanism or mechanisms of selectivity through which recognition accuracy may be influenced by instructional sets. Sets of 4 stimuli (pictures) each were displayed before, after, or before and after the tachistoscopic presentation of a single critical picture. The Ss had to report which 1 of the 4 alternatives was used as the critical picture. The stimuli within each set of 4 alternatives were either highly similar to one another or distinctively different from one another. The results indicated that perceptual processes were capable of being selectively tuned. This result stands in contrast to some previous research which indicated that recognition accuracy was influenced solely by memorial processes.

Since at least the time of Külpe's (1904) experiments on abstraction, psychologists have been concerned with the mechanism or mechanisms of selectivity through which accuracy of recognition is influenced by instructional sets. The existence of several thorough reviews of this general topic (e.g., Egeth, 1967; Pierce, 1963) permits us to single out for careful examination one of the most important and compelling arguments concerning the locus of such a mechanism.

Lawrence and Coles (1954) attempted to assess the relative importance of three possible selectivity mechanisms that might be operative in a perceptual recognition experiment: perceptual selectivity, selective remembering, and modification of response availability. Previously, investigators interested in examining selective processes in recognition had relied on a Before vs. After technique, i.e., one in which a to-be-reported aspect of the display was specified either before or shortly after the tachistoscopic presentation of a test stimulus. Accuracy of report was usually found to be higher with instructions preceding the stimulus display (e.g., Boring, 1924; Chapman, 1932), and several investigators had concluded that this result was an indication that *S*s perception of the stimulus could be selectively directed to critical aspects of the stimulus that had been specified by the instructions. However, Lawrence and Coles realized that this apparent selectivity might be reflecting a memory process rather than a perceptual one. Specifically, they reasoned that the Before instructions might simply have been able to modify the memory trace more quickly than the After instructions since there was

---

[1] This research was supported by the Advanced Research Projects Agency, Department of Defense, and was monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1235 with the Human Performance Center, University of Michigan. This research was carried out while the first author was on a National Science Foundation Fellowship. The authors would like to make a permanent record of their gratitude to Gail Egeth and Gail August for their contributions.

no delay in receiving the former kind of instructions although there was for the latter. In order to circumvent this difficulty, Lawrence and Coles resorted to a more elaborate experimental design. They reasoned that a percept must be richer and more full of detail than a memory trace. Granting this, consider the possible effect of giving alternatives (four verbal labels, one of which was appropriate to a briefly presented test picture) either before or after the presentation of the test picture. If the objects connoted by the four verbal labels are quite disparate, then even the relatively crude memory trace ought to carry enough information to enable *S*s to discriminate fairly adequately among them and thus choose the correct alternative. Hence, there should not be much decrement if the alternatives which connote disparate objects are presented after rather than before the test picture. However, if the objects connoted by the alternatives are similar to one another, then the crude memory trace may well be inadequate for discrimination when the alternatives are presented after the picture; but if the alternatives are presented before the picture then *S*s may be able to selectively direct their attention to those aspects of the picture that can best distinguish among the four alternatives. Therefore, in their experiment the existence of perceptual selectivity would be evidenced by an interaction between kind of alternatives, i.e., similar (the objects connoted by the verbal alternatives were similar to one another) or dissimilar (the objects connoted were different from one another) and time of presentation of the alternatives, i.e., before vs. after the test picture. In particular, perceptual selectivity ought to result in less difference in recognition accuracy between similar and dissimilar alternatives when the sets of alternatives are presented before rather than after the test picture. Since the predicted interaction was not statistically significant those authors concluded that the presentation of alternatives influenced either memory or response availability rather than perception. Their conclusion was greatly strengthened by the further finding that time of presentation of alternatives failed to produce even a significant main effect.

Although the logic underlying the design of their experiment still seems sound, we propose that the experiment itself does not provide an opportunity for a mechanism of perceptual selectivity to operate. Consider a hypothetical case where the four similar alternatives "church," "school," "house," and "barn" are presented before a test picture. In order for perceptual selectivity to be manifested *S* would first have to consider the alternatives and determine a set of critical features which could distinguish among the alternatives. He would then have to direct his attention to those parts of the test picture that contain these critical features. However, it is not clear just how *S*s could extract a set of critical visual features from a set of four verbal labels since it is not at all obvious just what set of features would distinguish reliably among the four *classes* of objects named above. In short, it does not seem reasonable to expect a perceptual process to be modified by this kind of conceptual information. Consideration of this point leads to a revision of the Lawrence and Coles experiment which does seem to provide a fair test of the perceptual-selectivity hypothesis. This revision involves the use of pictorial alternatives as well as pictorial test stimuli. In this situation *S* would have an opportunity to examine a set of alternatives in order to find features which distinguish among them, and he could then selectively direct his attention to the informative areas of the test picture.

METHOD

Subjects.—Forty-eight male and 48 female students at the University of Michigan served as paid volunteers and were tested individually.

Apparatus.—The test pictures were displayed for 20 msec. in a Gerbrands two-field mirror tachistoscope. In order to reduce the persistence of afterimages, the adaptation field consisted of a jumble of small bits and pieces of pictures. A small black fixation cross was clearly visible in the adaptation field at the viewing distance of 23 in. At that viewing distance a typical test picture subtended about 7° of visual angle. However, since the pictures did not have regular outlines, such a determination of an average size is only an approximation.

Procedure.—In an attempt to remain fairly close to the methodology of the Lawrence and Coles (1954) study, the present experiment was originally designed as a 2×2 factorial design with alternatives presented either before or after the test picture, and with the alternatives within a set being either similar to or dissimilar from one another. However, the results from a group of pilot *S*s indicated that the Before conditions imposed much greater memory loads on *S*s than did the After conditions. This point is treated further in the Results and Discussion sections; for now it is only important to note that as a result of the pilot data the experimental procedure was modified. Another level was added to the time-of-presentation-of-alternatives factor, viz., Before *and* After (Before-After). With this design, the chief comparison is between the After conditions and the new Before-After conditions. (For a precedent in the use of this design see Long, Reid, & Henneman, 1960.) Although the logic of the original Lawrence and Coles study is still intact, in the present experiment it is in the Before-After conditions that the effects of perceptual selectivity might be expected to occur unobscured by the confounding effects of a difficult memory task.

A different set of eight randomly chosen men and eight randomly chosen women served in each of the six experimental conditions. The *S*s were briefly familiarized with the nature of tachistoscopically presented stimuli by requiring them to describe a picture shown for 20 msec. They were given five such trials with the same picture, after which the picture was exposed for several seconds. Then, *S*s were given instructions appropriate to the conditions to which they had been assigned and these instructions were followed by five practice trials which were also appropriate to their assigned conditions.

In the Before-After conditions, on each trial *S*s were shown a plaque containing four pictures which they were allowed to examine for 30 sec. At the end of 30 sec. *E* said "Time" which was the cue for *S*s to look into the tachistoscope and fixate the black cross. As soon as they had done so they were to say "Ready" at which signal *E* flashed the picture for 20 msec. in the center of the field. The *S*s then looked back at the plaque containing the alternatives and pointed to the one that they thought had been flashed. They were permitted up to 30 sec. to reach a decision, and they were instructed to respond even if they were not certain of their answer. At the end of the 30-sec. decision period another set of four alternatives was presented and the procedure was repeated. There were 20 trials not including the five given for practice. The *S*s were paid $1.00 for participating, and, in addition, received 5¢ for every correct choice.

The Before conditions were essentially the same as the above with the single difference that the alternatives were removed while *S*s were peering into the tachistoscope. Therefore they were available before the presentation but not after. The *S*s indicated their choice by pointing to the appropriate position on a blank plaque that was divided into four positions by two perpendicular lines. This plaque was kept in place for 30 sec. so that the interstimulus intervals would be identical across conditions.

In the After conditions, *S*s were told to wait until they were informed that it was time for them to look into the tachistoscope; this waiting period was 30 sec. While *S*s were looking into the apparatus a plaque containing four alternatives was slipped into place. Immediately after the presentation of the stimulus picture *S*s looked down at the plaque and then made their choice.

Stimuli.—The sets of pictures used in the Similar conditions were constructed by searching through magazines and merchandise catalogs for instances in which there were four similar pictures of essentially the same object. (For example, pictures of four different styles of men's shoes were used.) Care was taken to select pictures of roughly the same size for a given set of alternatives. The alternatives within these sets of similar alternatives were at least as similar to one another as the objects connoted by the similar alternatives in the Lawrence and Coles study, since in the present experiment all four visual alternatives would be given the same label (e.g., shoe), while in the Lawrence and Coles study the similar alternatives were disparate enough to be associated with different labels.

The selected pictures were cut out and several monochromatic reproductions of each were made on a Thermofax copier. The 25 clearest four-picture sets were used; 20 in the experimental series and 5 in the practice series. From each set one of the alternatives was chosen to be used as a test picture and a copy of each of these pictures was reserved for this purpose. Then, copies of the four alternative stimuli in each set were attached to cardboard plaques so that they could be displayed as sets. The pictures were arranged on the plaque as if at the corners of a square. Consistent use of this arrangement allowed for identification of the critical member (i.e., the set member identical to the test picture) by a simple pointing response. This was particularly useful in the After conditions since the alternatives themselves were not present when *S* was making his decision. Each of the four possible positions contained a critical member equally often in the experimental series.

The sets of pictures for the Dissimilar conditions were constructed by recombining the pictures used in the Similar conditions. From a particular set of four similar alternatives only one was included in any given set of dissimilar alternatives. Again, the sets of alternatives were mounted on plaques at the corners of an imaginary square. The 20 test stimuli of the Similar series were also the test stimuli in the Dissimilar series. The critical members corresponding to these test pictures were kept in the same relative positions on their plaques in the two series. For example, if a particular shoe served as a test picture in the Similar series and the critical set member corresponding to it was displayed in the upper left of its plaque, then this same shoe would be a test picture in the Dissimilar series and would again correspond to a set member in the upper left of its plaque. To carry this illustration further, it should be made clear that this shoe would be presented as an alternative along with three other shoes in the Similar series, but in the Dissimilar series it would be displayed along with pictures of three quite disparate objects, e.g., a motorcycle, a radio, and a vacuum cleaner.

Six different orders of presentation of the stimuli were used in each condition during the course of the experiment.

RESULTS

Each *S* received a score from 0 to 20 based on the number of trials on which his choice was correct; these data were used in the analysis below.

Pilot *S*s had indicated that the Before conditions were qualitatively different from the other conditions because of the burden placed upon them of having to remember the positions in which the various alternatives had appeared on their plaques. Although this may not be an extremely difficult task in itself, it should be recalled that the 30-sec. examination period was supposed to be spent in a search for critical features, not in paired-associate learning. Thus the Before conditions not only imposed greater memory loads on *S*s but may have also reduced the time period during which the first stage of selectivity (determination of a critical set of discriminating features) operates. This preliminary finding led to the inclusion of the Before-After conditions in the experimental design. Now, the logical test of the perceptual selectivity hypothesis consists in a comparison of the After conditions with the Before-After conditions. Although the Before conditions were retained in the design and their results reported below, they did not enter into the statistical analysis.

Two steps were taken to reduce the error variance of the data. First, sex was used as an additional factor in the

TABLE 1

M and *SD* of Number of Correct Responses on 20 Trials as a Function of Time of Presentation of Alternatives, Similarity of Alternatives, and Sex of *S*

	Time of Presentation of Alternatives					
	Before		After		Before and After	
	M	*SD*	*M*	*SD*	*M*	*SD*
Males						
Similar	7.75	1.71	8.00	3.71	11.00	3.68
Dissimilar	13.75	3.34	16.25 (15.12)[a]	1.64 (2.31)[a]	15.12	2.87
Females						
Similar	7.50	2.69	4.75	1.98	9.00	3.12
Dissimilar	12.88	3.39	13.37 (12.62)[a]	2.94 (4.23)[a]	14.75	2.49

[a] These values are based on the data obtained before deviant *S*s were replaced.

analysis of variance.[2] Second, *S*s were eliminated from consideration if their scores deviated by more than 2 *SD*s in either direction from the means of their respective cells. The *S*s eliminated in this way (only two) were replaced with new *S*s. The means and standard deviations for all conditions are presented in Table 1.

An analysis of variance performed on the data of the After and Before-After groups indicated the following significant effects: Time of presentation of alternatives, $F\ (1, 56) = 5.91$, $p < .05$; similarity of alternatives, $F\ (1, 56) = 75.16$, $p < .01$; sex of *S*, $F\ (1, 56) = 7.59$, $p < .01$. The only significant interaction was the one of major interest—that between time of presentation of alternatives and similarity of alternatives, $F\ (1, 56) = 5.15$, $p < .05$. Examination of the data in Table 1 indicates that the nature of this interaction is exactly as expected on the perceptual-selectivity hypothesis, viz., there is less difference in recognition accuracy due to the similarity of alternatives under the Before-After conditions than under the After conditions.

Examination of Table 1 also indicates that recognition accuracy is lower in the Before conditions than in the Before-After conditions in all four possible comparisons. This supports the assumption that *S*s in the Before conditions had greater memory loads and/or less time to search for critical features.

Discussion

The major findings of the present experiment are: (*a*) Presentation of pictorial alternatives both before and after a test picture results in greater recognition accuracy than the presentation of alternatives only after the test picture, and (*b*) time of presentation of the alternatives interacts with the similarity of the alternatives in just the way that is predicted on the assumption of a mechanism of perceptual selectivity. Both of these results lend support to the notion that perception can be selectively directed by instructional information, and both stand in direct contradiction to the results of the Lawrence and Coles study. The major reason for this discrepancy in results was advanced in the introduction, viz., the conceptual information used to specify the alternatives in the Lawrence and Coles study (i.e., the verbal labels) was not sufficient to modify the perceptual recognition process. The results from the present Before conditions also suggest that any perceptual selectivity that might have been operative in the Before conditions of the Lawrence and Coles study might have been obscured by the increased memory requirements of this kind of condition.

Although an interpretation of the results of the present experiment in terms of perceptual selectivity seemed quite justified, it was pointed out to the authors

---

[2] Thanks are due to Robert G. Pachella, who ran all of the *S*s in the present experiment, for bringing our attention to the performance difference between the male and female *S*s.

that it is possible to account for these results solely in terms of some well-established principles of short-term memory.[3] Assume that perception of a test picture is uninfluenced by the presentation of alternatives. In order to respond correctly, *S*s must compare a memory trace of the test picture with some or all of the four alternatives. It seems reasonable to think that *S*s would "like" to make these comparisons in an intelligent order; specifically, it seems plausible to think that they would like to compare it with the most likely alternative first. Note, however, that *S*s in the After conditions cannot know which alternative is most likely because they do not get to see the alternatives until after the test picture is flashed. But *S*s in the Before-After conditions could make use of this strategy since they have at least some knowledge of the alternatives and their positions, and thus could first compare their memory trace of the test picture with the alternative that they consider most likely to yield a match. If *S*s in the Before-After conditions used this strategy successfully on some of the trials then on the average they could reach a decision on the basis of fewer comparisons than their counterparts in the After conditions. Since each comparison involves the inspection of an alternative, and since every inspected alternative is a potential source of interference with the memory trace, fewer comparisons result in less interference. Therefore the Before-After conditions should result in greater recognition accuracy than the After conditions. Furthermore, the magnitude of interference effects in short-term memory varies directly with the degree of similarity between the interfering events, so the Similar conditions ought to result in lower overall recognition accuracy than the Dissimilar conditions. It is also possible to predict the obtained interaction between similarity and time of presentation of alternatives since similarity of alternatives has less opportunity to affect accuracy differentially in the Before-After conditions than in the After conditions, because the former require fewer comparisons with the potentially interfering alternatives. (Note that this argument assumes that paired-associate learning of pictorial alternatives and their relative positions occurs during the first, i.e., "Before," segment of a trial in the Before-After conditions, just as was the case in the Before conditions.)

As a test of the validity of the preceding argument an additional eight males and eight females were run in a Before-After condition with Similar alternatives that was modified in such a way that the alternatives were presented in different locations on their plaques before and after the test picture. This change made it impossible for *S*s to utilize knowledge of the positions of the alternatives and thus seek out first the alternative they felt was most likely. In all other respects it was identical to the corresponding condition in the main experiment. The results: male *M*, 11.35 correct; female *M*, 9.12 correct. A comparison of these figures with the corresponding ones from Table 1 indicates that these *S*s performed no worse in the modified situation than the original *S*s performed in the unmodified task, and therefore that the proposed memory effect in the Before-After conditions was probably unimportant. Furthermore, since this result indicates that *S*s in the Before-After conditions did *not* memorize and utilize the positions of the alternatives and since previously presented results indicate that *S*s in the Before condition *did* (and in fact had to) memorize these positions, the assumption that the Before-After conditions are more sensitive to perceptual selectivity than the Before conditions is further strengthened.

Further justification for omitting the Before conditions from consideration because of the large memory component in that task is given by the results of a recent study by Steffy and Eriksen (1965). Either before or after the brief presentation of a test stimulus (a single nonsense form) a set of three alternative nonsense forms was presented, also very briefly. There was a variable delay of from 10 to 700 msec. between the two presentations (the duration of the delay is the interval between the offset of one display and the onset of the other). Thus, except for the great difference in durations, this study bears an obvious formal similarity to the present one. Steffy and Eriksen found that presentation of the alternatives after the test stimulus resulted in higher recognition accuracy than presentation of the alternatives before the test stimulus. Equally important, as the interval between the two presentations increased from 10 to 700 msec., accuracy decreased in the conditions in which the set of alternatives was presented before the single test stimulus, but in the conditions in which the test stimulus appeared first, accuracy remained at a constant level. It was concluded that the set of three alternatives was harder to remember over the delay interval than the single test stimulus because of the greater interference engendered by the larger number of items (see, e.g., Melton, 1963). It seems reasonable to invoke the same principles to account for the superiority of the Before-After conditions to the Before conditions in the present study.

---

[3] The authors would like to take this opportunity to thank Arthur W. Melton for his advice in the execution of this experiment, and in particular for this ingenious interpretation of the results.

In light of the overall pattern of results presented above it was concluded that selective visual perception could indeed be demonstrated in the laboratory. These results seem to be consistent with those of Harris and Haber (1963) on the effects of various coding strategies on recognition accuracy for tachistoscopically presented stimuli. They found that *S*s were capable of encoding the various aspects of a multidimensional stimulus roughly in the order of importance of those dimensions. Such behavior is clearly adaptive since the visual image resulting from a complex tachistoscopic presentation fades rapidly. Therefore, if the valuable dimensions are encoded (i.e., prepared for storage in memory) early, there is a greater probability that subsequent responses to them will be correct. The application of this general rationale to the present study is straightforward; *S*s in the Before-After conditions had an opportunity to determine which dimensions were likely to be effective for discriminating within a set of alternatives, and they also had the opportunity to use this knowledge in extracting information from the test picture. The *S*s in the After conditions could not know what kind of information to extract first; they got what they could from the flash, but there was an excellent probability that, in the Similar conditions, it would not be adequate for discrimination. Hence the interaction between similarity and time of presentation.

REFERENCES

Boring, E. G. Attributes and sensation. *Amer. J. Psychol.*, 1924, **35**, 301–304.

Chapman, D. W. Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, 1932, **44**, 163–174.

Egeth, H. Selective attention. *Psychol. Bull.*, 1967, **67**, 41–57.

Harris, C. S., & Haber, R. N. Selective attention and coding in visual perception. *J. exp. Psychol.*, 1963, **65**, 328–333.

Külpe, O. Versuche uber Abstraktion. *Ber. u. Kongr. F. exp. Psychol.*, 1904, 56–68.

Lawrence, D. H., & Coles, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **47**, 208–214.

Long, E. R., Reid, L. S., & Henneman, R. H. An experimental analysis of set: Variables influencing the identification of ambiguous, visual stimulus-objects. *Amer. J. Psychol.*, 1960, **73**, 553–562.

Melton, A. W. Implications of short-term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1–21.

Pierce, J. Determinants of threshold for form. *Psychol. Bull.*, 1963, **60**, 391–407.

Steffy, R. A., & Eriksen, C. W. Short-term, perceptual-recognition memory for tachistoscopically presented nonsense forms. *J. exp. Psychol.*, 1965, **70**, 277–283.

(Received August 8, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 550–555

CUE UTILIZATION IN SERIAL LEARNING[1]

SAM C. BROWN AND EUGENE D. RUBIN[2]

Kansas State University

4 groups of 30 Ss each learned either a 12-item serial list of low-meaningful (Lo-M) or high-meaningful (Hi-M) trigrams, with or without differently colored backgrounds surrounding each of the trigrams. After serial learning (SL), the list was broken into successive pairs for paired-associate (PA) learning; the odd-numbered items (trigrams, colors, or both) becoming stimuli and the even-numbered trigrams their responses. The results showed relatively faster PA learning for color than trigram stimuli in middle serial positions under the Lo-M list, but not under the Hi-M list. Under no conditions did color facilitate SL. The results were interpreted as indicating stronger color-trigram associations in middle positions and/or more interference from color in end positions. Relevance of the results for sequential and positional associations in SL is also discussed.

Research concerned with cue utilization in verbal learning has been largely restricted to the paired-associate (PA) paradigm. The research reported here was designed to investigate cue utilization in serial learning (SL), and in particular, to determine the differential use of added color cues in different serial positions of the list.

The procedure used in PA studies of cue utilization consists of two parts. First, *S* is required to learn a list of pairs comprised of compound stimuli, such as trigrams on colored backgrounds (e.g., Weiss & Margolius, 1954). Then a test is made of the associative strength of each stimulus component by means of a transfer task, which involves the separate presentation of each stimulus element alone for response-term recall. The present research used a comparable procedure. During SL the primary trigram members of the serial list each were presented on a differently colored background, thereby providing *S* with a potential extra stimulus cue (color) for each succeeding trigram in the list. After SL a transfer test was used to evaluate the relative associative strength of immediately adjacent color-trigram and trigram-trigram associations in each serial position of the list.

Despite a paucity of empirical data on cue utilization in SL, it seems reasonable to suppose that such utilization will depend both upon (*a*) the distinctiveness of the extra stimulus cues relative to the primary serial members of the list and (*b*) the serial position within the list. First, greater cue utilization should be expected for more distinctive added cues, as shown in PA tasks for cues of lower formal intralist similarity (Cohen & Musgrave, 1966; Newman & Taylor, 1963) and of higher meaningfulness (Sundland & Wickens, 1962; Underwood, Ham, & Ekstrand, 1962). Second, since difficulty of SL varies systematically with serial position, the utilization of added cues as functional stimuli should be greater in the more difficult middle than easier and more discriminable end positions of the serial list (e.g., Murdock, 1960).

---

[1] This investigation was supported by Public Health Service Research Grants MH 10019 and MH 11179 from the National Institute of Mental Health. The authors also wish to thank Alfred Volkuwitz for his assistance in the first experiment and Robert C. Haygood and William F. Battig for their valuable suggestions in the preparation of the manuscript.

[2] Now at the University of Michigan.

Evidence for a relatively greater utilization of added serial cues in middle than end positions in a serial list previously has been reported by Kausler and Trapp (1963). The authors' interpretation was that the added cues (2-digit numbers), although producing a significant overall decrement in SL, served to vary the distinctiveness of the primary serial members (low-association value trigrams) within the list. However, Kausler and Trapp did not evaluate sequential associative connections between extra stimulus cues and the primary trigrams of the list. Thus it is not presently known whether added cues develop functional stimulus properties during SL.

To help clarify the picture, this research compared SL and subsequent PA transfer with and without color, and under different levels of trigram meaningfulness.

METHOD

Subjects.—The 120 paid *S*s were all summer-session students at Kansas State University. Assignment of *S*s to groups was based on a predetermined order which assured that the groups were filled at the same rate. Each *S* served for a single session of 2 hr. or less.

Serial learning.—All *S*s learned a 12-item serial list of CVCs to a criterion of one errorless trial. The 60 *S*s in the Trigram (T) group were shown all trigrams of the list on a uniformly white background. For the 60 *S*s in the Trigram plus Color (TC) group, each trigram of the list appeared on a differently colored background. The same serial order of trigrams and colors was used on each trial. Equal numbers of *S*s (30) within both the T and TC group learned a low-meaningful (Lo-M) list and a high-meaningful (Hi-M) list, thereby yielding a total of four SL conditions. Learning was carried out under standard anticipation procedure, with an asterisk as a cue for the first item in the list. The serial list members, each contained on separate 3 × 5 in. index cards, were presented manually through a card-display box at a 3-sec. presentation rate with a 6-sec. intertrial interval. An audible click from an electrically operated timer was used to pace list presentation.

All *S*s were given typical instructions for SL. In addition, instructions to *S*s in the TC condition mentioned that color would surround each trigram, but emphasized that the task was to learn the order of the trigrams, not the order of the colors. No mention was made of the subsequent PA transfer task.

PA transfer task.—Two min. after completion of SL, *S*s (10) within each TC condition learned a PA list under one of three different stimulus conditions. All three six-pair lists contained the six original (even-numbered) serial trigrams (without their colored backgrounds) as response terms. However, stimulus-term composition for each response term was either a compound of the immediately preceding (odd-numbered) serial trigram and its background color (TC); the preceding serial trigram alone (T); or the preceding serial background color alone (C). Each of the above three stimulus lists also was learned by a third of the *S*s in the T serial condition. For all lists, the order of pair presentation was different on every trial.

Assignment of *S*s within each SL condition to the three PA stimulus conditions was made on the basis that total SL trials to criterion and variability of these scores be matched, so as to equate for initial learning ability. Learning of the PA lists was carried out under the recall (study-test) procedure to a criterion of one errorless trial or a maximum of six trials. The cards containing the six pairs were presented at a 3-sec. rate with a 12-sec. interval between study-test series, and trials. Each *S* was told about the composition of the stimulus term, but not the rules of pairing. During both PA learning and SL, all *S*s were required to pronounce their responses.

Before leaving the laboratory, *S*s were tested for defects in color vision, using the Pseudo-Isochromatic Plates produced by the American Optical Company. Four color-deficient *S*s were identified and their data discarded. Data from four replacement *S*s were substituted in the analyses.

Materials.—The 12 Lo-M and 12 Hi-M trigrams were of 40–45% and 99–100%

Archer (1960) association value, respectively. The lists were selected so as to minimize formal intralist similarity both within the Lo-M and Hi-M list and similarity of meaning within the Hi-M list. The 12 colored backgrounds, selected in part for relative ease of naming, were as follows: YELLOW, BLACK, ORANGE, GREEN, PINK, GRAY, PURPLE, RED, BLUE, OLIVE, WHITE, and BROWN. During SL, each of three different trigram orders for the Lo-M and the Hi-M list was learned by an equal number of *S*s (10) in Cond. T. The *S*s in Cond. TC learned the same trigram orders, each combined with a different serial order of color. The colored backgrounds, made of construction paper, were pasted on the index cards for SL and PA learning. When combined with a trigram, the color frame completely surrounded the verbal unit.

RESULTS

Serial learning.—Table 1 shows the mean number of total errors (failures to respond correctly) for each of the four SL conditions. As can be seen, the effect of color varied depending upon whether the list was Lo M or Hi M. For the Lo-M list, more errors were made under Cond. TC than T; for the Hi-M list, performance was practically identical for the two conditions, leading to a significant Hi-Lo × Color interaction, F (1, 112) = 3.98, $p < .05$. Significantly fewer errors also occurred under the Hi-M (54.1) than Lo-M list (84.4), F (1, 112) = 23.70, $p < .001$, but the overall superiority of T over TC fell short of significance ($p > .10$). However, pairwise comparisons showed Cond. TC and T to differ significantly from one another on the Lo-M list, F (1, 56) = 4.78, $p < .05$.

TABLE 1

MEAN NUMBER OF ERRORS ON SL AND PA LEARNING

SL		PA Transfer			
		TC	T	C	Total
Color					
Easy	52.9	4.6	9.6	3.4	5.9
Hard	95.7	8.6	9.2	7.1	8.3
Total	74.3	6.6	9.4	5.3	7.1
No Color					
Easy	55.2	6.0	7.4	1.6	5.0
Hard	73.3	8.4	8.4	2.4	6.4
Total	64.2	7.2	7.9	2.0	5.7
Totals		6.9	8.7	3.7	

The learning curves showed comparable rates of acquisition for T and TC conditions, and examination of serial-position effects failed to reveal differential utilization of color in different positions of the list. Besides an overall significant bowed serial-position effect, F (11, 1,283) = 61.26, $p < .001$, only the Hi-Lo × Position interaction proved reliable, F (11, 1,283) = 2.98, $p < .001$, reflecting a flattened serial-position curve for the Hi-M as compared with the Lo-M list. However, even this difference was eliminated when the curves were plotted in percentage of error scores at each serial position. Comparison of fast and slow learners and analyses of total number of errors made before and after the first correct response in each serial position, revealed results basically the same as those reported above for total errors.

PA transfer.—The mean number of total errors made under the three stimulus conditions is presented in Table 1, separately for each SL condition. In all instances performance was best for the C condition, next best for TC and poorest for T, except for identical TC and T scores under the Lo-M T-serial condition. Analysis of covariance (using unitized deviation, z, SL scores for each S as the covariate) showed the overall stimulus effect to be highly significant, F (2, 107) = 10.41, $p < .001$. Further analyses revealed performance for the C condition to be significantly better than for the combined TC and T conditions, F (1,

107) = 18.36, $p < .001$, which in turn did not differ significantly from one another, F (1, 107) = 2.45, $p > .10$. Although all interactions involving Stimulus conditions with Color and/or Hi-Lo M lists were nonsignificant (p's $> .20$), average performance under the C condition was significantly poorer when preceded by SL with (5.3) than without color (2.0), F (1, 107) = 4.22; $p < .05$. In agreement with the SL results, there were significantly fewer overall PA errors for the Hi-M (5.4) than Lo-M list (7.4), F (1, 107) = 4.41, $p < .05$, and the difference favoring T over TC conditions was not significant ($p > .10$). However, the Hi-Lo × Color interaction was eliminated for the PA task ($F < 1$).

To determine the extent to which color was used as a stimulus in different parts of the PA list, several comparisons were made between PA stimulus conditions as to relative performance on pairs from beginning, middle, and end positions of the serial list. Of particular interest is the comparison of individual pair performance for the T and C Stimulus conditions for *S*s who first learned the Lo-M TC-serial list. Figure 1 shows the mean number of total PA errors under these two stimulus conditions for each of the six pairs, arranged along the abscissa according to their original positions in the serial list. As shown, when the stimuli were trigrams (T), more PA errors were made on responses previously from the middle than ends of the serial list; when the stimuli were colors (C), better performance was displayed for the middle- than end-position responses. Both the overall T-C × Position interaction, F (5, 270) = 2.82, $p < .025$, and the quadratic component thereof, F (1, 54) = 7.74, $p < .01$, were significant by extended-trend analysis of variance (Grant, 1956). Figure 1 does not include the PA transfer curve for Cond. TC, which showed better performance in Pair Positions 1 and 4, and nearly equivalent performance across the remaining four pairs. For all remaining PA conditions, the T and TC curves showed a close resemblance to the bow-shaped curve depicted in Fig. 1 for the T condition, while the transfer curve for the C condition did not differ appreciably from a horizontal straight line, showing nearly perfect performance in each position.

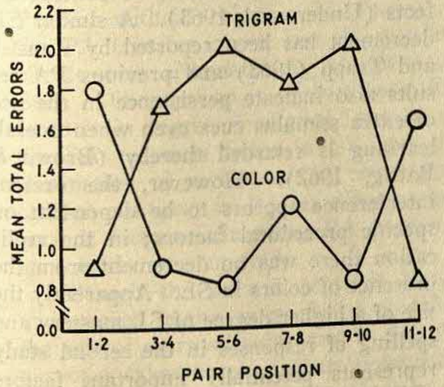


FIG. 1. Mean total PA errors at each pair position for the Trigram and Color Stimulus conditions under the Lo-M list.

Replication.—Because of the unique nature of the color data, replication seemed desirable so as to place greater confidence in the results. Reported here are the partial findings from a second study by Read (1966), which included the Lo-M SL and PA conditions as used in the first experiment. The *S*s were 54 regular session college students, divided equally into T and TC serial groups. In all important respects the procedure was the same as that reported above, except for (*a*) the use of different and slightly lower 31–35% Archer (1960) association value trigrams; (*b*) the requirement that *S*s spell their responses; and (*c*)

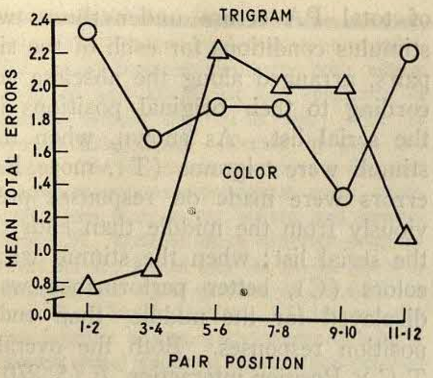


FIG. 2. Mean total PA errors at each pair position for the Trigram and Color Stimulus conditions in the replication.

learning of the serial and PA lists to a criterion of two consecutive errorless trials.

Figure 2 shows the PA transfer curves for the T and C Stimulus conditions for the TC serial group. Apart from an elevation of the C curve, the two curves are highly similar to those depicted in Fig. 1. Extended-trend analysis of variance again revealed the quadratic component of the T-C × Position interaction significant, F (1, 48) = 5.94, $p < .025$. The PA transfer curve for Cond. TC resembled that of Cond. T. Other findings paralleled closely those of the first experiment, except for virtually equivalent mean total errors for T (151.0) and TC (151.7) groups during SL, and failure to find a significant PA stimulus effect ($p > .20$) between Cond. TC (9.9), T (8.6), and C (9.2). However, performance under Cond. C was still numerically better for *S*s who had not seen color on SL (7.1) as compared with those who had (11.3), although the difference was nonsignificant ($p > .20$).

DISCUSSION

The crucial finding of this study is that the serial effect of prior SL is not only canceled but actually reversed when the added cues of SL serve as stimulus terms in PA learning. This provides strong evidence for the hypothesis that added color cues in SL are utilized more in the middle (more difficult) portion of the list than at the ends. It should be noted, however, that an alternative explanation is possible; some aspect of the added cues may be causing greater interference at the ends of the list than in the middle. For example, if *S*s are forming stronger associations between colors and their embedded trigrams at the ends of the list, presentation of an end color during PA learning may evoke greater response competition against the required response of the trigram which followed it in SL. A rather elaborate experiment is needed to clarify this ambiguity thoroughly. Despite the ambiguity of interpretation, however, it is clear that added color cues exert an effect, and that their effect depends rather heavily on the part of the serial list under consideration.

Nonetheless, the presence of both trigrams and color in the first experiment apparently served to produce a significant source of interference under the Lo-M list condition during SL, possibly because of *S*s' efforts to learn both elements in the serial associative process. Under the conditions of the present research it was not possible for *S*s to ignore the primary (trigram) members of the serial list, as is possible in PA studies of context effects (Underwood, 1963). A similar SL decrement has been reported by Kausler and Trapp (1963) and previous PA results also indicate persistence in the use of extra stimulus cues even when overall learning is retarded thereby (Brown & Battig, 1962). However, the present interference appears to be dependent on specific procedural factors; in the replication there was no decrement from the presence of colors in SL. Apparently the use of a higher degree of SL mastery and spelling of responses in the second study represents potentially important factors determining the overall effect of color in SL.

The failure of the present experiments

to show differential cue utilization during SL conflicts with the Kausler and Trapp (1963) results. The use of a longer serial list (12) as compared with the Kausler and Trapp Study (8), plus other possible procedural differences, may account for at least part of the discrepancy. The fact that no significant effect due to color was observed under the Hi-M SL condition is congruent with the failure also to find intraserial effects during PA learning, and may mean that these *S*s relied primarily on the Hi-M (word) element of the compound as the functional stimulus in SL.

The present results also appear relevant for recent studies using the SL to PA paradigm as a means of studying the role of sequential and positional associations in SL (e.g., Young, in press). Specifically, there are indications that prior experience with items in SL inhibits subsequent PA learning. This was evidenced by the finding in both experiments of slower color-trigram PA learning under conditions of SL with color as compared to SL without color. The implication is that previous failures to show significant positive transfer from SL to PA learning (e.g., Young, 1959; in press) may not necessarily reflect the absence of sequential associations in SL, but rather the cancellation of positive transfer by interference from paired adjacent-serial items. Along the same lines, it is interesting to note that *S*s learning the color-trigram PA list without color in SL showed a flat PA transfer curve, i.e., SL position effects were eliminated in PA learning. Such a finding argues against both differences in degree of original learning of the PA response terms in SL and the elimination of serial-position cues (e.g., ordinal position) in PA learning, as important factors influencing PA performance in the present study.

Although the present research leaves unanswered many questions concerning the role of extra stimulus cues in SL, the approach taken herein appears potentially useful for the continued investigation of intralist learning processes, and in particular, for the study of functional and nominal stimuli in sequential learning tasks.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Brown, S. C., & Battig, W. F. Partial serial-position constancy in nonserial paired-associate learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 42–47.

Cohen, J. C., & Musgrave, B. S. Effects of formal similarity on cue selection in verbal paired-associate learning. *J. exp. Psychol.*, 1966, **71**, 829-838.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Kausler, D. H., & Trapp, E. P. Irrelevant cues and serial learning effects. *Psychol. Rep.*, 1963, **12**, 798.

Murdock, B. B. The distinctiveness of stimuli. *Psychol. Rev.*, 1960, **67**, 16–31.

Newman, S. E., & Taylor, L. R. Context effects in paired-associate learning as a function of element-sharing among stimulus terms. *J. verbal Learn. verbal Behav.*, 1963, **1**, 243–249.

Read, J. D. Extraserial cues in verbal learning. Unpublished Masters thesis, Kansas State University, 1966.

Sundland, D. M., & Wickens, D. D. Context factors in paired-associate learning and recall. *J. exp. Psychol.*, 1962, **63**, 302–306.

Underwood, B. J. Stimulus selection in verbal learning. In C. H. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 33–48.

Underwood, B. J., Ham, M., & Ekstrand, B. Cue selection in paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 405–409.

Weiss, W., & Margolius, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48**, 318–322.

Young, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 544–559.

Young, R. K. Serial learning. In D. L. Horton and T. R. Dixon (Eds.), *Verbal behavior theory and its relation to general S-R behavior theory.* Englewood Cliffs, N. J. Prentice-Hall, in press.

(Received August 8, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 556–561

SHORT-TERM MEMORY, WORD CLASS, AND SEQUENCE OF ITEMS[1]

HENRY LOESS

College of Wooster

120 Ss received 24 word triads in a Peterson-type short-term memory test. Items contained high or low frequency words, were homogeneous or heterogeneous with respect to taxonomic category, and successive tests of a given category were massed or distributed over the trials. When massed, retention was maximum for the first item of a category, significantly poorer in most cases for the second and third; when distributed, no loss occurred for retention of words, a slight loss for intra-item position. Heterogeneous items showed maximum PI after 1 trial. Word frequency did not interact with the other variables. The interpretation is that PI does not develop across taxonomic classes and that within classes it dissipates if items are sufficiently spaced.

Several studies have reported a rapid development of proactive interference (PI) when testing the short-term retention of individual items. Specifically, PI has been shown to be a function of length of the retention interval (Keppel & Underwood, 1962), time between presentations (Loess & Waugh, in press; Peterson & Gentile, 1965), and class membership or similarity of items (Loess, in press; Wickens, Born, & Allen, 1963). Typically, these studies have employed a modified Peterson and Peterson (1959) technique whereby an item, such as a consonant trigram, is presented briefly, followed by a task designed to prevent rehearsal until the cue to recall.

In the Wickens et al. study, *S*s were presented a series of consonant trigrams (CCCs) or 3-digit numbers (NNNs) and then a test item for which half the *S*s were shifted to the alternate class (NNN or CCC). Marked interference was found when proactive and test items were from the same class, but no evidence of PI was observed when proactive and test items were from different classes of materials. In the Loess (in press) study, *S*s were presented a series of word triads in which the words came from a single taxonomic category such as trees, birds, or countries. A rapid buildup of PI was observed as a function of number of prior items. Then *S*s received a series of items with words from a second category. Retention of the first showed no evidence of PI, but PI developed rapidly as additional items were presented.

The present study was designed to determine the importance of preexperimental dependencies when items are organized by taxonomic category and different sequences are presented across trials. Specifically, *S*s received a series of triads consisting of words of either high or low frequencies of occurrence according to the norms of Cohen, Bousfield, and Whitmarsh (1957). Retention of each item was measured after 9 sec. Major interest centered on the buildup of PI as a function of item composition and number of prior items.

---

[1] This work was supported by Research Grant No. MH 08537-02 from the National Institutes of Health to the College of Wooster. Lee DeCoster and Keith McCartney helped in the collection and analysis of the data.

METHOD

Subjects.—The *S*s were 120 paid volunteers from psychology classes at the College of Wooster divided randomly into six groups of 20. Most had served in previous verbal-learning studies but not with materials of the present study.

Materials.—The items were word triads prepared from lists taken from the Bousfield, Cohen, and Whitmarsh (1958) study on associative clustering in recall. The words represented eight categories: animals, vegetables, boys' names, professions, birds, countries, cloths, and musical instruments. They were originally obtained from the norms of Cohen et al. (1957) and consisted of separate lists of 10 high frequency and 10 low frequency words from each category. For the present study a random procedure was used to eliminate 1 word from each list and arrange the remaining 9 into groups of 3's. In this manner, three triads of high frequency words and three of low frequency words were constructed for each of the eight categories. These (homogeneous) items were used in four of six conditions. For the remaining conditions two sets of 24 heterogeneous items were composed from the separate pools of words used in Cond. 1–4. Items were constructed by a random procedure so that each item contained words from three categories.

Conditions.—There were six conditions: Cond. 1—On Trials 1–8 a homogeneous, high frequency item was presented randomly from each of the eight categories; likewise on Trials 9–16 and 17–24. Cond. 2—Same as above with low frequency items. Cond. 3—All three homogeneous, high frequency items were presented from one randomly chosen category followed by items of a second category, and so forth, for all eight categories. Cond. 4—Same as above with low frequency items. Cond. 5—All 24 heterogeneous, high frequency items were presented in random order. Cond. 6—Same as above with low frequency items.

Procedure.—The *S*s were run individually, each receiving a different random order of items which met the requirements of the condition. The items were presented on 2 × 2-in. slides by means of a Kodak Carousel projector, and intervals were controlled by Hunter 111C times. Each slide was presented for 1.5 sec. and it took an additional .5 sec. for the tray to rotate to the next slide. When the trials began, the word "Ready" appeared on a screen. This was followed by a triad which *S* spoke aloud. Next in succession appeared five slides each containing a 6-digit number and *S* was instructed to read each number as rapidly as possible. Then, a slide containing the word "Recall" signaled the beginning of a 10.5-sec. recall interval. A 6-sec. rest preceded the ready signal for the next trial. Thus, each item was exposed for 1.5 sec., the retention interval was 9 sec., and the total time between successive items was 30 sec., except for an additional 60 sec. between Trials 12 and 13 to change Carousel trays. Results were scored in terms of number of correct items—all words recalled in correct intra-item order, and number of correct words—without regard for order.

RESULTS

Analyses of variance, described below, indicated that the main effects and all interactions involving word frequency were statistically unreliable, therefore the data for high and low frequency conditions were combined in Fig. 1 and 2. Figure 1A shows that recall of words was maximal throughout when items were presented in random order for blocks of eight trials (Cond. 1 and 2, combined). Figure 1B shows that when the same items were presented category by category retention decreased consistently for the second and third items of each category and rose to a maximum on the first trial of each new category (Cond. 3 and

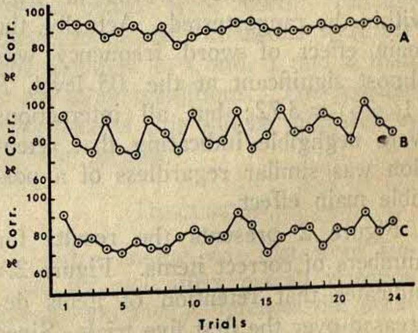


FIG. 1. Recall of words as a function of item composition and sequence of presentation.

4, combined). Figure 1C, representing the recall of words arranged into items without regard for category membership (Cond. 5 and 6, combined), shows a large decrement from Trial 1 to Trial 2 and smaller decrements over the next four trials. The increase observed on Trial 13 may reflect the effects of the extra-long rest interval following Trial 12.

An analysis of variance of a 2 × 2 × 3 design, comparing the overall retention of homogeneous items (Cond. 1–4), revealed that Sequence (random vs. ordered) was significant, *F* (1, 76) = 4.31, as were Number of Prior Items of the same category (0, 1, or 2), *F* (2, 152) = 30.08, and Sequence × Prior Item interaction, *F* (2, 152) = 17.10. The significant interaction reflects the decreasing recall that occurred over items of specific category when presentations were ordered but not when they were presented randomly (Fig. 1B vs. Fig. 1A). It indicates that the development of PI is limited to items from the same category and is dependent upon the time relations between successive presentations from the same category.

The failure of word frequency (high vs. low) main effects or interactions to be significant indicates that PI development is the same for high and low frequency words at least within the range tested. Actually, the main effect of word frequency was almost significant at the .05 level, *F* (1, 76) = 3.72; but all interactions were negligible indicating that retention was similar regardless of a possible main effect.

Figure 2 presents the results for numbers of correct items. Figure 2A indicates that retention of items decreased over the first five trials. Since Fig. 1A shows that word recall remained maximal, the decreases observed in Fig. 2A largely reflect transposition errors (incorrect intraitem recall). Figure 2B shows the same general trend as Fig. 1B. Analysis of variance over the above conditions produced a significant Sequence × Prior Item interaction, *F* (2, 152) = 13.61, as it had in the previous analysis. Again, neither the main effect nor any interaction involving word frequency was significant.

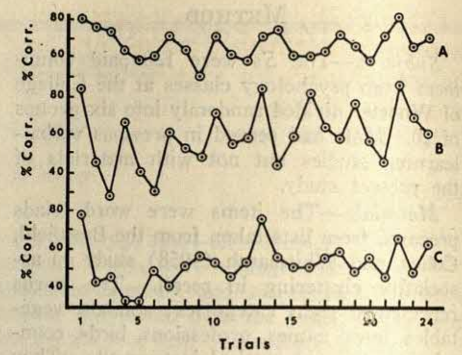


FIG. 2. Recall of intact items as a function of item composition and sequence of presentation.

Figure 2C shows the same trend as Fig. 1C. Recall of items dropped sharply from Trial 1 to Trial 2, rose from Trial 12 to Trial 13, and dropped again on Trial 14.

Since each data point in Fig. 2 represents a proportion of individuals who responded correctly, trial-by-trial differences were tested by means of *z* scores for correlated measures. For Fig. 2A, comparisons with Trial 1 showed a significant decrease between Trial 1 and Trial 4 ($z = 2.00$) suggesting a small but significant accumulation of PI over trials. Comparisons of performance on all *adjacent* trials revealed a significant change from T 9 to T 10 and from T 22 to T 23 ($p <$ .05 in both cases), but no reason is evident, other than chance, to account for these significant effects.

For Fig. 2B, there was a significant decrease in retention from the first to the second items for four of the eight categories (T 1 vs. T 2, T 4 vs. T 5, T 13 vs. T 14, and T 19 vs. T 20) ($p < .05$); and from the first to the third items in six of the eight categories. In all seven instances there was a significant increase from the third item of one category to the first item of the next ($p < .05$).

For Fig. 2C, comparisons of adjacent trials showed four significant differences (T 1 vs. T 2, T 12 vs. T 13, T 13 vs. T 14, and T 21 vs. T 22) ($p < .05$). Although it is difficult to interpret the meaning of these results, three of the four differences were expected a priori, and there is no reason to expect the fourth is other than a chance effect. Specifically, the T 1 vs. T 2, T 12 vs. T 13 and T 13 vs. T 14 differences are consistent with the findings of Peterson and Gentile (1965) who found PI to develop rapidly when trials were massed but to dissipate over a 91-sec. interval.

Table 1 presents the relative proportions of different types of errors. It shows that transpositions and intrusions were relatively more frequent with high frequency words, omissions more common with low frequency words, and transpositions more common with homogeneous than with heterogeneous items. These results are consistent with expectations that more intraitem interference should occur with high than with low frequency words and with homogeneous than with heterogeneous items.

TABLE 1

Total Number and Relative Proportions of Different Kinds of Errors

Cond.	Description	Transpositions	Omissions	Intrusions	Total No. Errors
1	Hom-Hi-R	.62	.22	.16	278
2	Hom-Lo-R	.51	.41	.08	374
3	Hom-Hi-O	.52	.29	.19	390
4	Hom-Lo-O	.47	.36	.17	454
5	Het-Hi-R	.39	.43	.18	435
6	Het-Lo-R	.34	.51	.15	471

Note.—Items were either homogeneous or heterogeneous, of high or low frequency of occurrence, and presentation was either random or organized according to taxonomic category.

TABLE 2

Total Number and Relative Proportions of Different Kinds of Intrusions

Cond.	Prev. Item	Other Items	Extra-Exp't'l	Approp. Categ.	Total No. Intrusions
1	.02	.57	.41	.90	46
2	.00	.39	.61	.87	31
3	.73	.15	.12	1.00	74
4	.65	.19	.16	.91	76
5	.59	.25	.16	.54	79
6	.68	.20	.12	.50	72

Table 2 shows that for Cond. 1 and 2 the previous item was nonfunctional as a source of overt interference. For these conditions such items always represented a different category than the current item. On the other hand, for Cond. 3 and 4 the previous item was the major source of intrusions. Virtually all such intrusions occurred on 1 of the 16 trials when the previous item represented the current-item category. For conditions using homogeneous items almost all assignable intrusions were from the appropriate category, compared to .52 when heterogeneous items were used. Finally, extraexperimental intrusions were far more common in Cond. 1 and 2 than for the other conditions. The possible importance of this will be discussed in the next section.

Discussion

A major finding was that word frequency did not affect overall retention although expected differences did occur in the frequencies of different kinds of errors. Since word frequency was controlled over trials and categories, the successive development and release of PI from one category to the next under massed conditions cannot be attributed to differential word availability or preexperimental associations between words and the category names. Neither can preexperimental associations be used to explain the dissipation of PI under conditions of distributed presentations of items within categories.

The present study is consistent with studies in which intertrial interval (ITI) has been the experimental variable. As the time between successive items (X_1 and X_2) is increased, PI is consistently found to decrease. Loess and Waugh (in press) found no PI for intervals beyond 120 sec., and Peterson and Gentile (1965) found no PI for the first item of a block of items when blocks were separated by 91 sec. In the present experiment the same is true. Within a category, when X_1 and X_2 occur on adjacent trials, X_1 exerts a considerable PI on the recall of X_2 as indicated by the incidence of overt intrusions. When X_1 and X_2 are separated in time by presentations from other categories, X_1 does not exert measurable PI for the recall of X_2. Neither do the intervening items.

If the PI effect of X_1 is reduced by retroactive effects occurring during an extended X_1-X_2 interval (filled or unfilled), such events should also act proactively on the recall of X_2. If, on the other hand, PI from X_1 dissipates from the effects of prior items acting during the X_1-X_2 interval then, as Peterson and Gentile (1965) have pointed out, such interference should also act proactively on the recall of X_2. A third alternative is simply that PI dissipates in time.

In the present study, .14 of all intrusions were from extraexperimental sources when items of a category were massed, and .51 when they were distributed. In comparison, Loess and Waugh (in press) found .16 extraexperimental intrusions with a 0-sec. ITI and .74 with a 3-min. ITI, the latter being the approximate time taken to present the intervening items in the present study. Such data suggest that extraexperimental interference replaces intraexperimental interference as ITI increases. Whether this reflects an increase in the strength of the former and a decrease in the strength of the latter, or a decrease in the latter while the former remains constant, is not clear. In the first instance, the total amount of interference should remain the same, but in the second instance it would not; and this should lead to better retention with distributed trials accompanied by a smaller and smaller proportion of intraexperimental intrusions as ITI is increased, precisely what occurred in the present study. An alternative explanation would be that briefly exposed items are subject both to decay and to interference (Peterson & Gentile, 1965). In this case, extraexperimental intrusions would simply reflect the absence of intraexperimental interference at the time of recall.

In conclusion, the cause of PI dissipation remains in doubt; but the present results extend previous findings indicating that PI does not develop from one class of materials to another. Assuming the same holds true for RI (retroactive interference), the use of dissimilar materials during retention intervals of STM studies is supported. Finally, taxonomic category and intertrial interval might well be utilized more often in STM research to overcome problems of confounding in the use of repeated measures (Keppel, 1965; Peterson, 1965).

REFERENCES

Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychol. Rep.*, 1958, **4**, 39–44.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957 ONR Contract Nonr- 631 (00), University of Connecticut.

Keppel, G. Problems of method in the study of short-term memory. *Psychol. Bull.*, 1965, **63**, 1–13.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *J. verbal Learn. verbal Behav.,* 1962, **1,** 153–161.

Loess, H. B. Short-term memory and item similarity. *J. verbal Learn. verbal Behav.,* in press.

Loess, H., & Waugh, N. C. Short-term memory and inter-trial interval. *J. verbal Learn. verbal Behav.,* in press.

Peterson, L. R. A note on repeated measures in the study of short-term memory. *Psychol. Bull.,* 1965, **64,** 151–152.

Peterson, L. R., & Gentile, A. Proactive interference as a function of time between tests. *J. exp. Psychol.,* 1965, **70,** 473–478.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.,* 1959, **58,** 193–198.

Wickens, D. D., Born, D. G., & Allen, C. K. Proactive inhibition and item similarity in short-term memory. *J. verbal Learn. verbal Behav.,* 1963, **2,** 440–445.

(Received August 8, 1966)

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

The titles and authors of accepted papers are listed here. In 1965 this program was being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology and the outcome of this trial is presentely being evaluated.

This listing represents the entire backlog of manuscripts accepted by this journal. They appear here 2 or 3 months after acceptance due to the printing lag. Such listing is intended to allow readers to become aware of research many months in advance of journal publication.

A Model of Learning for Simple Repeating Binary Patterns: Paul C. Vitz* and Thomas C. Todd: Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Generalization Gradients as Indicants of Learning and Retention of a Recognition Task: Harry P. Bahrick,* Sandra Clark, and Phyllis Bahrick: Department of Psychology, Ohio Wesleyan University, Delaware, Ohio 43015.

Effects of Response-Induced Stimulus Change on Human Discrimination: F. Robert Treichler,* Barbara Hann, and Sally J. Way: Department of Psychology, Psychological Clinic, 221 Kent Hall, Kent State University, Kent, Ohio 44240.

Unlearning and Competition in List-1 Recall: Theresa S. Howe*: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California 94720.

An Evaluation of Subjective Probability in a Visual Discrimination Task: William C. Howell*: Department of Psychology, Human Performance Center, Ohio State University, 404-B West Seventeenth Avenue, Columbus, Ohio 43210.

Negative Contrast in Human Probability Learning as a Function of Incentive Magnitudes: John A. Schnorr and Jerome L. Myers*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts.

Image Size and Instructions in the Perception of Depth: Albert J. Dinnerstein*: New York Medical College, Department of Physical Medicine and Rehabilitation, Welfare Island, New York, New York 10017.

Compatibility and Time-Sharing in Serial Reaction-Time: Steven W. Keele*: Department of Psychology, University of Oregon, Eugene, Oregon 97403.

Is a Test Trial a Training Trial in Free Recall Learning?: Roy Lachman* and Kenneth R. Laughery: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

Effects of Feedback and Instructional Set on the Control of Cardiac-Rate Variability: Peter J. Lang,* L. Alan Sroufe, and James E. Hastings: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.

(Continued on page 585)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 562–573

STRENGTH OF ALTERNATIVE RESPONSES AND SUBSEQUENT CHOICES[1]

WILLIAM B. PAVLIK AND ROBERT LEHR

Rutgers University

In 3 experiments, the number of reinforced and/or nonreinforced runs to the 2 arms of a T maze were manipulated by means of forcing procedures. Choice behavior was studied during subsequent extinction trials, which were all free choice. The results indicate: (a) choice is a function of the relative strengths of the alternative responses, (b) choice is relatively insensitive to number of reinforcements, (c) choice is substantially correlated with earlier speed differences, and (d) the usual PRE for choice behavior does not appear when experience with the alternative responses is controlled.

Many investigators have studied extinction in choice situations as functions of incentive variables manipulated during acquisition. It frequently happens, however, that choice behavior *during the acquisition phase* varies systematically as a function of the manipulated incentive variable. When such acquisition differences have occurred, then differences during the subsequent extinction period may not be attributed unequivocally to the operation of the incentive variable; the latter will have been confounded with differential experiences with the alternative responses during the acquisition phase. Recently, for example, it has been shown that a conventional partial-reinforcement extinction effect (PRE) occurs in a choice situation when continuous (CRF) and partial reinforcement (PR) groups are allowed to differ during training, but that a reversed PRE occurs when the groups do not differ during training in terms of relative numbers of experiences with the correct and incorrect alternatives (Pavlik & Reynolds, 1964; Spear & Pavlik, 1966).

Findings such as a reversed PRE in a choice situation not only point out the importance of confounds between independent variables and experiences with the alternative responses, but they also appear to require interpretation (Pavlik & Reynolds, 1964) in terms of a long-standing theoretical notion (e.g., Luce, 1959; Spence, 1936), which is that choice behavior reflects less the absolute strengths of the responses involved than it does the relative strengths of the responses involved. Thus, a variable, such as PR, that leads to a large number of nonreinforced incorrect responses during training (compared to a CRF control group) may lead to great resistance to extinction of the correct response—not because the correct response is strong, but because the competing incorrect response has been rendered very weak. When the strength of the competing incorrect response is made equal to that operating in the CRF control group, then the PR group no longer shows superior resistance to extinction.

Although the notion of relative response strengths determining choice behavior has been stressed theoretically, only a few systematic studies using infrahuman *S*s have submitted this notion to test. Hill, Cotton, and Clayton (1963) gave rats a fixed number of reinforced runs to the correct arm of

---

[1] This research was supported by United States Public Health Service Grant M-07844.

a T maze while varying the number of reinforced and/or nonreinforced runs to the incorrect arm. When choice behavior was assessed in a subsequent series of free extinction trials, choice of the correct arm both decreased as the number of prior reinforcements in the incorrect arm increased, and increased as the number of prior nonreinforcements of the incorrect arm increased. One surprising finding was that, when only reinforcements had been experienced in the two arms, then subsequent choice behavior was approximately at chance. This occurred even when twice as many reinforcements had been taken in the correct arm (16) as in the incorrect arm (8). This latter result is different from that of an earlier study by Ramond (1954), who gave free and forced trials such that one alternative was reinforced twice as often as the other. Ramond's *S*s consistently preferred the more frequently reinforced alternative. In a study which deliberately confounded the number of experiences with the two response alternatives, Lachman (1961) showed that rats given 75% of their initial runs to the nonreinforced incorrect arm chose the correct arm more frequently on subsequent free-choice trials than did rats given either 50% or 25% of their runs to the incorrect arm, a result similar to that reported by others (e.g., Birch, 1955; Denny & Dunham, 1951; Uhl, 1966).

The experiments reported here were designed to shed further light on the effects upon choice behavior during extinction (where the reinforcement contingencies for both responses are the same) of systematic variations in the strengths of one or both of the responses. The T maze was used because the total set of choice responses consists of only two mutually exclusive members. By using forcing procedures (i.e., lowering a door at the choice point so that the rat may enter only one of the arms of the maze), it was possible to directly manipulate the number of times each response occurred as well as the reinforcement contingencies operating on each response occurrence. It was assumed that the strength of each of the two responses could be independently manipulated. This assumption required the further, perhaps unwarranted, assumption that generalization between the two arms and/or responses was negligible.

In addition to choice behavior, records were kept of response speed in various portions of the apparatus. These speeds are of interest because they have been often regarded (e.g., Hull, 1952; Spence, 1956) as direct measures of the absolute strengths of the individual responses. As such, they provided an alternative method of assessing the relative strengths of competing responses.

EXPERIMENT I

Experiment I was designed to provide training experiences such that (*a*) the strength of the correct response would be constant across experimental groups, and (*b*) the strength of the incorrect response would be varied systematically across groups. This was accomplished by giving all groups the same number of reinforced runs to the correct side of the maze while varying the number and/or reinforcement contingencies of runs to the incorrect side. Specifically, all groups received 40 reinforced runs to the correct side and either 20 nonreinforced (Group 20N), 10 nonreinforced (10N), zero (0), 10 reinforced (10R), or 20 reinforced (20R) runs to the incorrect side. Following training, choice was studied during 30 extinction trials, none of which was forced.

Method

Subjects.—The *S*s were 50 naive male albino rats that weighed between 180 and 200 gm. at the start of the experiment. Forty rats were purchased from Sprague-Dawley, Incorporated, and 10 from Holtzman Company. The *S*s were randomly assigned to the five experimental groups with the restriction that eight Sprague-Dawley and two Holtzman rats be in each group.

Apparatus.—The apparatus, described in detail elsewhere (Pavlik & Reynolds, 1964), consisted of a single unit T maze painted flat black and covered with clear Plexiglas. Doors at the choice point and goal boxes could be used either to prevent entry into or retracing from various portions of the maze. Photoelectric circuitry provided measures of Start Speed, Choice Speed of either the correct or incorrect arm at the choice point, and Arm Speed in either the correct or incorrect arm.

Procedure.—Five days prior to the start of training, *S*s were placed on a feeding schedule of 1-hr. access to 10 gm. of Purina lab chow each day. Ad lib water was available in the home cage. This schedule was maintained throughout the experiment, the feeding period beginning approximately 15 min. after testing each day. On the final pretraining day, each *S* was given 5-min. habituation to the start box and stem of the maze.

Both arms of the maze were open on the first training trial for each *S*, and the reinforced choice made was designated that *S*'s correct side. All subsequent training trials were forced by closing one of the doors at the choice point. Training progressed for 5 days, on each of which each *S* received eight reinforced runs to the correct side, and either zero, two, or four runs to the incorrect side. The runs to the incorrect side were randomly interspersed among the runs to the correct side. For *S*s in Groups 10R and 20R, runs to the incorrect side were reinforced; for *S*s in Groups 10N and 20N, they were not reinforced. Group 0 *S*s, of course, made no runs to the incorrect side during training. Each reinforcement consisted of three 45-mg. Noyes pellets. Detention in the goal box was for 10 sec. on nonreinforced trials, and for either 10 sec. or until the food was eaten on reinforced trials. The intertrial interval was approximately 30 sec.

Following training, each *S* received 3 days of extinction, at 10 trials per day. The extinction trials were the same as the training trials except that both arms were open on each trial and no reinforcements were present. Noncorrection procedures were followed throughout.

Results

Acquisition.—A mean speed score (1/latency) was computed for each *S* for each day of acquisition training for each speed measure. No group differences occurred for start speed, correct choice speed, or correct arm speed; there was, however, a tendency for Groups 10R and 20R to be slower in correct choice speed (see Fig. 1). The absence of any significant group differences on these measures supports the assumption that the strength of the correct response was held constant across groups.

Significant group differences did occur in incorrect choice speed and in incorrect arm speed, $F\ (3, 36) = 8.92$ and 4.48, respectively, $p < .01$. These comparisons do not include Group 0, which had no experience with the incorrect arm. The group differences, which grew larger across training days, are in the expected direction. Figure 1 presents the incorrect choice speed

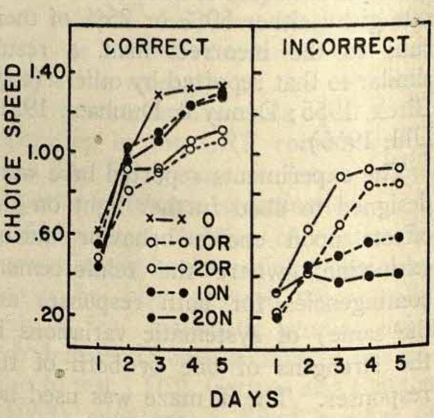


FIG. 1. Mean correct and incorrect choice speeds during acquisition training in Exp. I.

data; incorrect arm speed was quite similar.

Extinction speed data.—All speed measures declined significantly during extinction. No group differences appeared in the start-speed measure. A significant Groups × Days interaction occurred for correct choice speed, F (8, 90) = 2.25, $p < .05$. The form of the interaction was that Groups 20N, 10N, and 0 ran more rapidly early in extinction and also declined in speed more rapidly than did Groups 20R and 10R. The same Groups × Days pattern occurred for correct arm speed, but not at an acceptable level of reliability, F (8, 90) = 1.33.

For the incorrect choice speed measure, there was a significant Groups effect, F (4, 45) = 4.62, $p < .01$, with Groups 20R and 10R running faster throughout extinction than any of the remaining groups. The same pattern, faster running from Groups 20R and 10R, obtained for the incorrect arm speed measure, but not at a reliable level, F (4, 45) = 1.49.

Extinction choice data.—Mean number of choices of the correct arm during the 30 extinction trials is presented in Fig. 2. Significant effects of both Groups, F (4, 45) = 19.13, $p < .01$, and the Groups × Days interaction, F (8, 90) = 7.85, $p < .01$, were obtained. Although the ordering, by Groups, of the mean correct choices during extinction is directly related to the manipulated strength of the incorrect response, application of Tukey's tests indicates that Groups 20R and 10R do not differ at the .05 level of confidence; all other pairs of groups differ at at least the .05 level. The interaction of Groups with Days took the form of larger group differences early in extinction, with the magnitude of the differences, but not their direction, declining across days of extinction. Throughout the extinction period, the mean percentage of correct choices by Groups 20R and 10R did not deviate appreciably from chance, being below 60% for each block of 10 extinction trials.

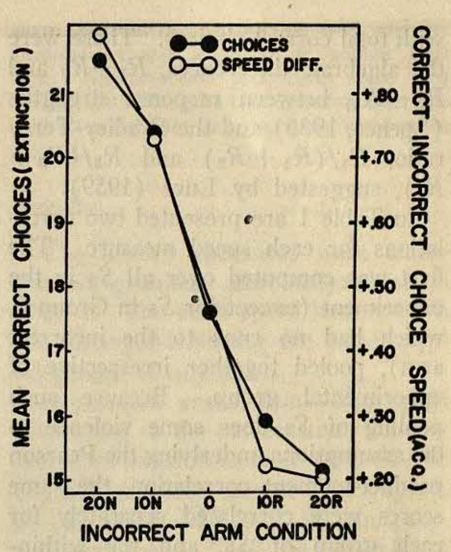


FIG. 2. Mean number of correct choices during extinction and mean algebraic difference between choice speeds on the last acquisition day of Exp. I. (The points were plotted to produce a visual "best fit" between choice and speed data.)

Correlations of choices with speeds.—As various theorists (Luce, 1959; Spence, 1936) have suggested that choice may be predicted from various combinations of measures of the absolute strengths of alternative responses, and since running speed has been considered to directly reflect the absolute strength of an instrumental response, correlations were computed between total correct choices during extinction and each S's mean score on the final day of acquisition training for correct choice speed (R_1), incorrect choice speed (R_2), correct arm speed (R_3), and incorrect arm speed (R_4). In addition to the four single speed measures, two different indexes of relative response strength were correlated with total correct choices. These were the algebraic differences, $R_1 - R_2$ and $R_3 - R_4$, between response strengths (Spence, 1936) and the Bradley-Terry ratio, $R_1/(R_1 + R_2)$ and $R_3/(R_3 + R_4)$, suggested by Luce (1959).

In Table 1 are presented two correlations for each speed measure. The first was computed over all *S*s in the experiment (except for *S*s in Group 0, which had no runs to the incorrect arm), pooled together irrespective of experimental group. Because such pooling of *S*s does some violence to the assumptions underlying the Pearson product-moment correlation, the same scores were correlated separately for each group of *S*s, and the within-groups correlations were averaged via a *z*′ transformation procedure (Edwards, 1960). The mean within-groups correlation is presented in parentheses in Table 1. It will be noted that the mean within-groups correlations generally are smaller in value but in the same direction as the correlations computed directly over all *S*s. The smaller values of the correlations computed within groups probably stem from the smaller within-groups variances, especially in the correct-choice measure.

TABLE 1

PRODUCT-MOMENT CORRELATIONS ACROSS GROUPS AND, WITHIN PARENTHESES, THE MEAN WITHIN-GROUP CORRELATIONS BETWEEN TOTAL CORRECT CHOICES DURING EXTINCTION AND VARIOUS SPEED MEASURES FROM THE FINAL DAY OF ACQUISITION

Speed Measure	Exp. I	Exp. II	Exp. III
R_1 Correct Choice	+.49*	+.30*	+.43*
	(+.24)	(+.09)	(+.48)
R_2 Incorrect Choice	−.70*	−.47*	−.76*
	(−.32)	(−.37)	(−.40)
$R_1 - R_2$	+.75*	+.53*	+.78*
	(+.34)	(+.33)	(+.59)
$R_1/(R_1+R_2)$	+.76*	+.35*	+.74*
	(+.42)	(−.14)	(+.50)
R_3 Correct Arm	.00	−.07	+.14
	(+.55)	(−.02)	(+.28)
R_4 Incorrect Arm	−.47*	−.36*	−.67*
	(+.22)	(−.20)	(−.18)
$R_3 - R_4$	+.53*	+.35*	+.66*
	(+.22)	(+.18)	(+.42)
$R_3/(R_3+R_4)$	+.51*	+.29*	+.67*
	(+.17)	(+.19)	(+.42)

* $p < .05$.

Generally, the choice-speed measures yielded higher correlations than did the arm speeds, and, among the single speed measures, speeds to the incorrect alternative yielded higher correlations than did speeds to the correct alternative. This is to be expected from the particular experimental manipulations employed. Finally, the very similar correlations yielded by the algebraic difference between speeds, $R_x - R_y$, and the Bradley-Terry ratio, $R_x/(R_x + R_y)$, leave little to choose between these models for predicting choice behavior. The close correspondence between mean correct choices and mean $R_1 - R_2$ is presented graphically in Fig. 2.

EXPERIMENT II

As the findings of Exp. I suggested that choice behavior during extinction is near chance when only reinforcements have been experienced following the alternative responses (i.e., Groups 20R and 10R), Exp. II was designed both to replicate certain of the findings of Exp. I and to more systematically explore the role, if any, of number of reinforcements in determining choice behavior. The reinforcement conditions were varied in both arms of the T maze in a 2 × 3 factorial arrangement. Groups of *S*s received either 8 reinforced or 8 nonreinforced runs to the incorrect arm, coupled with either 16, 32, or 64 reinforced runs to the correct arm. Group designations—64R-8R, 32R-8N, etc.— reflect the number of reinforced runs to the correct arm and whether the 8 runs to the incorrect arm were reinforced (R) or nonreinforced (N).

Method

Subjects.—The *S*s were 60 naive male albino rats, purchased from Sprague-Dawley, Incorporated, which weighed between 180 and 200 gm. at the start of the experiment.

Apparatus.—The apparatus was a single-unit T maze, consisting of a 12-in. start box, 12-in. stem, 22-in. arms, and 18-in. goal boxes. All portions were 4-in. wide and 4-in. high, painted flat black, and covered with clear Plexiglas. Plexiglas guillotine doors separated the start and goal boxes from the remainder of the apparatus; additional Plexiglas guillotine doors were located 1.5 in. from the choice point in either arm and were used to prevent entry into and/or retracing from either arm of the maze. The sole source of illumination in the black-walled experimental room was a 15-w. bulb suspended 40 in. above the choice point.

Photoelectric circuits provided for five measures of response speed. Start speed was measured from the opening of the start box door to the intersection of a beam 6 in. into the stem. Choice speed (to either arm) was measured over a 12-in. distance from the beam in the stem to a beam 6 in. past the choice point in either arm. Arm speed (in either arm) was measured over a 24-in. distance between beams located 6 in. past the choice point and 8 in. past the goal-box doors. The latter beams were located, in turn, 4 in. in front of the food cups, which were located 6 in. in front of the rear walls of the goal boxes.

Procedure.—Pretraining, deprivation schedule, magnitude of reinforcement, and the first acquisition trial procedures were the same as in Exp. I. Subsequent acquisition trials were all forced. Acquisition training proceeded for 8 days, on each of which *S*s in Groups 16R-8R and 16R-8N received two reinforced runs to the correct arm, *S*s in Groups 32R-8R and 32R-8N received four reinforced runs to the correct arm, and *S*s in Groups 64R-8R and 64R-8N received eight reinforced runs to the correct arm. During each of the last 4 days of acquisition training, each *S* received two runs to the incorrect arm which were randomly interspersed among his runs to the correct arm. Runs to the incorrect arm were reinforced for *S*s in Groups 16R-8R, 32R-8R, and 64R-8R, but were nonreinforced for *S*s in Groups 16R-8N, 32R-8N, and 64R-8N.

Following the completion of acquisition training, each *S* received 10 extinction trials on each of the following 2 days. Both arms were open on each extinction trial, and reinforcement was available in neither goal box. Noncorrection procedures were used throughout. The goal box detention intervals and intertrial intervals were the same as in Exp. I.

Results

The latency measures were transformed, grouped, and analyzed as in Exp. I. The start speed measure did not distinguish among the experimental groups at any point.

Acquisition speed data.—For speed of running to the correct arm, the correct arm reinforcements (64, 32, or 16) affected both correct choice speed, F (2, 54) = 4.52, $p < .05$, and correct arm speed, F (2, 54) = 15.50, $p < .01$, with 64 reinforcements yielding the fastest speeds and 16 reinforcements the slowest for both measures. The incorrect arm condition (8R or 8N) affected correct choice speed, F (1, 54) = 4.28, $p < .05$, with -8N groups faster than -8R groups, but

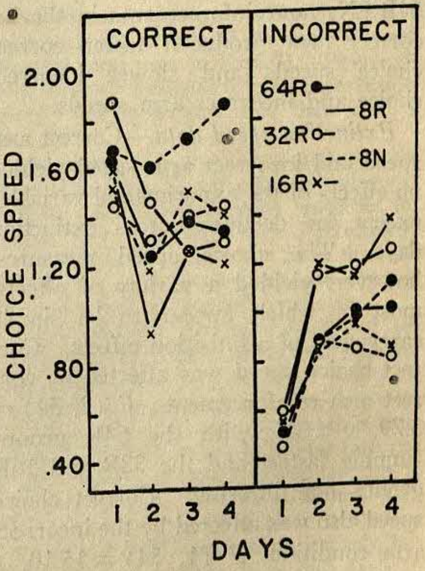


FIG. 3. Mean correct and incorrect choice speeds for the final 4 days of acquisition training in Exp. II.

did not affect correct arm speed. The interaction of the two conditions did not affect either correct choice or correct arm speed. The final 4 days of acquisition choice speed data are shown in Fig. 3. The reduction in correct choice speed is correlated with the introduction of runs to the incorrect arm.

On runs to the incorrect arm, the interaction of correct arm reinforcements with incorrect arm condition affected both incorrect choice speed, F (2, 54) = 7.74, $p < .01$, and incorrect arm speed, F (2, 54) = 5.85, $p < .05$. The form of the interaction is presented graphically in Fig. 3 for incorrect choice speed. Incorrect arm speed yielded the same pattern.

The general picture, then, is that number of reinforcements in the correct arm affected correct choice and correct arm speeds in the expected direction, but did not affect incorrect choice or incorrect arm speeds. With the exceptions of Groups 64R-8R and 64R-8N, nonreinforcements in the incorrect arm produced faster correct choice speeds and slower incorrect choice and incorrect arm speeds.

Extinction speed data.—Correct arm speed and incorrect arm speed yielded no effects of the experimental variables except for declines across extinction days. The choice speed measures, however, yielded a variety of effects, most of which appear to be simple carryovers of acquisition effects. Correct choice speed was affected by correct arm reinforcements, F (2, 54) = 3.79, $p < .05$, with the 64R groups running fastest and the 32R and 16R groups not differing. Correct choice speed also was affected by the incorrect arm condition, F (1, 54) = 15.10, $p < .01$, with the -8N groups running faster than the -8R groups; this difference diminished across extinction, producing an interaction with Days, F (1, 54) = 18.95, $p < .01$. Incorrect choice speed during extinction was not affected by correct arm reinforcements, but was affected by the incorrect arm condition, F (1, 54) = 17.61, $p < .01$, with the -8R groups running faster than the -8N groups; this difference also declined across extinction, generating an interaction with Days, F (1, 54) = 10.00, $p < .01$.

Extinction choice behavior.—The mean numbers of correct arm choices made during the 20 extinction trials are shown in Fig. 4. Analysis of these data yielded no significant effect of correct arm reinforcements. However, both the incorrect arm condition, F (1, 54) = 48.60, $p < .01$, and the interaction of the conditions in the two arms, F (2, 54) = 3.47, $p < .05$, yielded significant effects. The interaction (see Fig. 4) apparently stemmed from the smaller difference between Groups 32R-8R and 32R-8N than between

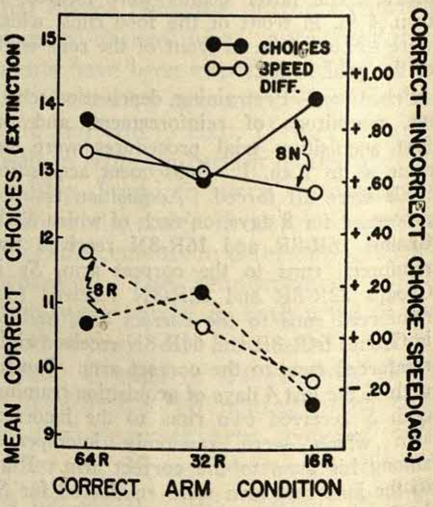


FIG. 4. Mean number of correct choices during extinction and mean algebraic difference between choice speeds on the last acquisition day of Exp. II. (The points were plotted to produce a visual "best fit" between choice and speed data.)

either of the remaining pairs of groups. Analyses of the comparisons involved in the interaction by both Tukey's tests and multiple Mann-Whitney tests revealed only that each -8N group made more correct choices than did its comparable -8R group; no reliable differences occurred either among -8R groups or among -8N groups. It also should be noted that the -8R groups performed at essentially chance (50%) levels throughout the extinction period. This confirms both the previous findings of Hill et al. (1963) and Exp. I.

Correlation of choices with speeds.—In Table 1 are presented correlations between total correct choices during extinction and various speed measures and combinations of speed measures from the final day of acquisition training. As in Exp. I, for each measure a correlation was computed across all *S*s irrespective of group, and a mean within-groups correlation was also computed. While these correlations are generally smaller than the comparable correlations from Exp. I, the same pattern emerged as in Exp. I. Generally, choice speeds yielded higher correlations than did arm speeds, and incorrect speeds yielded higher correlations than did correct speeds. Unlike Exp. I, the simple algebraic difference, $R_x - R_y$, between speeds tended to yield higher correlations than did the more involved Bradley-Terry ratio.

The general correspondence of group mean correct choices with group mean $R_1 - R_2$ is shown in Fig. 4. It will be noted that the correspondence was generally less close than in Exp. I, and that the mean $R_1 - R_2$ scores present a more consistent picture than do the mean correct choices.

Experiment III

As the results of Exp. I and II indicated that (*a*) nonreinforced experiences with the incorrect arm greatly enhanced subsequent choice of the correct arm, and (*b*) choice of the correct arm was relatively insensitive to the number of correct arm reinforcements, Exp. III was designed to explore the role of nonreinforcements in the correct arm.

A 3 × 2 factorial design was employed in which *S*s received either CRF or PR for runs to the correct arm, and were either reinforced for, nonreinforced for, or received no runs to the incorrect arm. Thus, the design consisted of three partial-reinforcement (CRF vs. PR) comparisons, one at each of three conditions of experience with the incorrect arm, and provides for replication of the previous findings of Pavlik and Reynolds (1964) and Spear and Pavlik (1966) who reported reversed PREs when experience with the two arms was held constant across groups.

Method

Subjects and apparatus.—The *S*s were 72 naive male albino rats of the same size and strain as were used in Exp. II. Twelve *S*s were randomly assigned to each experimental group. The apparatus was that used in Exp. II.

Procedure.—The procedures used were the same as in Exp. II with the following exceptions. Acquisition training proceeded for 4 days, on each of which each *S* received 10 runs to his correct arm. The *S*s in the -16R and -16N groups (see below) also received four daily runs to the incorrect arm.

Group 40C-16R *S*s were reinforced for each run to both the correct and incorrect arms. Group 40C-16N *S*s were reinforced for each run to the correct arm, but were nonreinforced for runs to the incorrect arm. Group 40C-0 *S*s were reinforced for each run to the correct arm, and received no runs to the incorrect arm. The *S*s in Groups 40P-16R, 40P-16N, and 40P-0 were treated as were *S*s in the 40C- groups, respectively, but were reinforced on only six of each day's 10 runs to the correct arm. The position of the runs to the incorrect arm in each daily block of trials and the reinforcement schedule for the 40P- *S*s were determined randomly with the restriction that, for all *S*s, the first and last run each day were to the correct arm and were reinforced.

Following acquisition training, each *S* received 10 extinction trials each day for 3 days. As in Exp. I and II, all extinction trials were free and reinforcement was available in neither goal box. Goal-box detention intervals and intertrial intervals were as in Exp. I and II.

Results

Acquisition speed data.—For correct choice speed, the interaction of correct arm schedule (40C- or 40P-) with incorrect arm condition (-16R, -16N, or 0) across Days was significant, F $(6, 198) = 9.00$, $p < .01$. The form of the interaction can be seen in Fig. 5. Schedule of reinforcements in the correct arm appeared to have affected correct choice speed only when the incorrect arm was nonreinforced. There were no comparable differences, however, in correct arm speed.

Both incorrect choice speed and incorrect arm speed were affected by the incorrect arm condition. (Groups 40C-0 and 40P-0 were, of course, not included in these analyses.) For incorrect choice speed, as shown in Fig. 5, the -16R groups both ran faster than the -16N groups, and this difference grew larger with training, F $(3, 132) = 9.28$, $p < .01$. The same effect of incorrect arm condition appeared in incorrect arm speed, but interacted with correct arm schedule and Days, F $(3, 132) = 7.33$, $p < .01$. The form of this interaction was that no speed differences occurred between the -16R groups across training, but Group 40C-16N ran progressively more slowly than did Group 40P-16N as training progressed. Thus, while nonreinforcements in the incorrect arm generally produced slower incorrect choice and arm speeds than did reinforcements in the incorrect arm, this effect was accentuated when only reinforcements had been experienced in the correct arm.

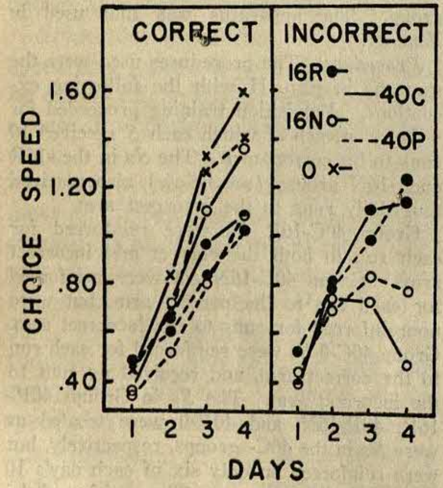


FIG. 5. Mean correct and incorrect choice speeds during acquisition training in Exp. III.

Relative speed of choice during training is presented in Fig. 6, on which are plotted the mean algebraic differences between correct and incorrect choice speeds on the final day of acquisition training. It will be noted that a large positive difference occurs only in the case of Group 40C-16N, which was the only group that experienced only reinforcements in the correct arm and only nonreinforcements in the incorrect arm. Group 40P-16R, conversely, chose the incorrect arm more rapidly than the correct arm; this group experienced nonreinforcements only in the correct arm.

Extinction speed data.—A variety of significant effects occurred for the extinction speed measures, but they may be summarized briefly. First, the correct arm schedule produced a significant conventional PRE for all speed measures, even those taken in the incorrect arm. This constitutes what has been called the "ubiquitous PRE" (Spear & Pavlik, 1966). Second, the PRE produced by the correct arm

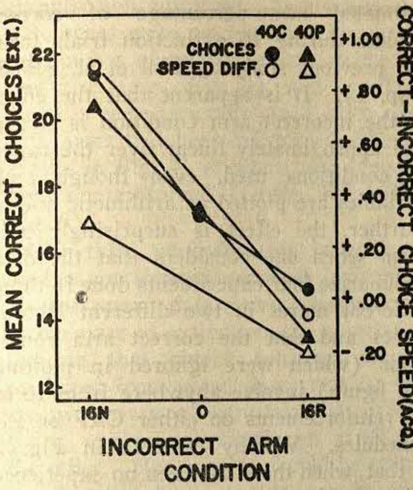


Fig. 6. Mean number of correct choices during extinction and mean algebraic difference between choice speeds on the last acquisition day of Exp. III. (The points were plotted to produce a visual "best fit" between choice and speed data.)

schedule interacted across Days with the incorrect arm condition in start speed, F (4, 132) = 24.50, $p < .01$, and in incorrect choice speed, F (4, 132) = 21.50, $p < .01$. In both cases, the interaction stemmed from a large PRE in the -16R comparison, a moderate PRE in the -0 groups comparison, and a small PRE in the -16N comparison. Third, the incorrect arm condition affected both correct arm speed and incorrect arm speed, in both cases the effect being that the -16R groups ran faster than the -0 or -16N groups, which did not differ.

Extinction choice behavior.—The mean number of correct arm choices made during the 30 extinction trials is shown in Fig. 6. There was again a strong effect of the incorrect arm condition, F (2, 66) = 45.43, $p < .01$, with the direction of the differences the same as in Exp. I and II. While there was a tendency toward a reversed PRE in both the -16R groups comparison and the -16N groups comparison, the absence of a comparable effect in the -0 groups comparison resulted in an overall effect of the correct arm schedule that did not attain an acceptable level of reliability, F (1, 66) = 2.00, $p > .05$. However, in no case was a PR group reliably more resistant to extinction of the correct choice response than was its comparable CRF group. Thus, there is only partial confirmation of the reversed PRE reported previously (Pavlik & Reynolds, 1964; Spear & Pavlik, 1966).

As in Exp. I, the group differences diminished as extinction progressed, all groups approaching chance levels of choice. For Group 40P-16R, choice of the correct arm was below chance on the first extinction day (32%) and increased toward chance levels as extinction progressed.

Correlations of choice with speeds.—As in Exp. I and II, correlations (both across all *S*s and within groups) were computed between the number of correct arm choices during extinction and various speed measures, and combinations thereof, taken on the final day of acquisition training. These are presented in Table 1. The correlations were of similar magnitude to those obtained in Exp. I. As in Exp. I, the two indexes of relative response strength, $R_x - R_y$ and $R_x/(R_x + R_y)$, yielded highly similar correlations. It is of interest to note that, although the correlation of individual correct choices with individual algebraic differences in choice speeds, $R_1 - R_2$, was as high in Exp. III as in Exp. I, there was a much closer correspondence of the group means for these variables in Exp. I than in Exp. III (see Fig. 2 and 6).

Discussion

The results of the three experiments described here may be summarized in terms of three generalizations. (*a*) The manipulation of the strength of the weaker (i.e., incorrect) response in a two-response choice situation has a large effect on subsequent free-choice behavior during extinction. (*b*) Choice behavior is relatively insensitive to number of reinforcements beyond some necessary minimum and is relatively indeterminant (i.e., near chance) when both response alternatives have been only reinforced. (*c*) Choice behavior is anticipated by, and is correlated substantially with relative speeds of the alternative responses when the correlation ignores the experimental grouping of *S*s.

The first generalization is shown graphically in Fig. 7, which shows the mean-percentage of correct choices during the first 20 extinction trials for the three experiments reported here and the estimated mean-percentage of correct choices during 16 extinction trials from the previous study by Hill et al. (1963, Exp. 1). It is apparent that the effect of the incorrect arm condition is strong and approximately linear over the range of conditions used, even though both variables are plotted as arithmetic scales. Further, the effect is surprisingly uniform when one considers that the data summarize four experiments done in three different mazes in two different laboratories and that the correct arm conditions (which were ignored in plotting the figure) involve anywhere from 16 to 64 reinforcements on either CRF or PR schedules. Worthy of note in Fig. 7 is that, when there has been no experience with the incorrect arm during training (the 0 condition), performance during extinction is generally above chance; this indicates that some discriminative behavior may be evidenced even though no "errors" have been made in training, a point made in quite a different context by Terrace (1963).

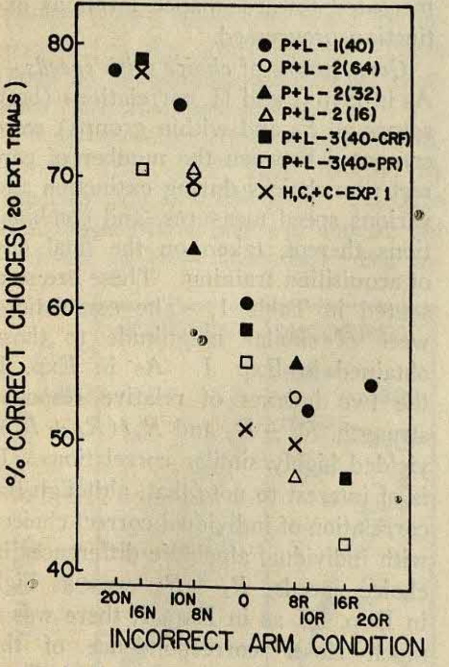


FIG. 7. Mean percentage of correct choices during the first 20 extinction trials for each of the three experiments reported here and for the first 16 trials from an earlier study by Hill, Cotton, and Clayton (1963, Exp. I). (P & L in the legend refers to the present experiments and H, C, & C to the Hill et al. experiment.)

The second generalization also is clearly shown in Fig. 7. At any single incorrect arm condition, the points plotted reflect anywhere from 16 to 64 correct arm reinforcements, and yet there are neither large nor consistent differences among the points plotted at any incorrect arm condition. Similarly, when incorrect arm behavior has been reinforced (all the points to the right of the 0 condition in Fig. 7), subsequent extinction performance is near chance, even though up to eight times as many reinforcements have been experienced in the correct arm as in the incorrect arm. Both findings suggest a relative insensitivity of choice behavior to number of reinforcements beyond some minimum.

Indeed, insofar as extinction choice behavior is anticipated by speed differences during acquisition, the second generalization can be extended to include such acquisition speed differences. Reference to Fig. 2, 4, and 6 indicates that large differences in speeds occurred only when the correct arm was consistently reinforced and the incorrect arm consistently nonreinforced. Significant speed differences were not produced by varying the number of reinforcements in the correct arm (e.g., Fig. 4), and speed differences were generally small when either both arms had been reinforced (the -8R groups in Fig. 4) or when both arms had been nonreinforced (Group 40P-16N in Fig. 6).

The third generalization is evidenced by the correlations reported in Table 1. As the speed data were collected prior to the start of extinction, the correlations indicate that choice behavior may be predicted from speed data; that is, since the two measures were taken at different times, one is not simply an artifact of the other. However, it should be pointed out that the choice speeds may be somewhat contaminated by *S*'s tendency to turn in one direction rather than the other. As *S*, on a forced trial, could not see which choice-point door was lowered until after the choice speed measure had commenced, a tendency to turn toward the "closed" arm on that trial would have resulted in a longer choice latency than would a tendency to turn toward the "open" arm. That something like this was happening is suggested by the fact that, for all three experiments, there were substantial correlations with the correct choice speed measure, but only moderate correlations with correct arm speed; the latter measure would have been free of any such contamination.

Finally, note should be taken of the fact that there was no evidence of a conventional PRE for choice behavior in Exp. III. In two of the three CRF vs. PR comparisons, the CRF group was (numerically) more resistant to extinction of the correct response than was the PR group. This appears to strengthen the conclusion, drawn previously (Spear & Pavlik, 1966), that resistance to extinction of a choice response is greater with higher percentages of reinforcement when experience with the alternative responses is not confounded with the independent variable.

REFERENCES

Birch, D. Discrimination learning as a function of the ratio of nonreinforced to reinforced trials. *J. comp. physiol. Psychol.*, 1955, **48**, 371–374.

Denny, M. R., & Dunham, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in the simple T maze. *J. exp. Psychol.*, 1951, **41**, 382–389.

Edwards, A. L. *Experimental design in psychological research.* Rev. ed. New York: Rinehart, 1960.

Hill, W. F., Cotton, J. W., & Clayton, K. N. Effect of rewarded and nonrewarded incorrect trials on T maze learning. *J. comp. physiol. Psychol.*, 1963, **56**, 489–496.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Lachman, R. The influence of thirst and schedules of reinforcement-nonreinforcement ratios upon brightness discrimination. *J. exp. Psychol.*, 1961, **62**, 80–87.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Pavlik, W. B., & Reynolds, W. F. Conventional and reversed partial reinforcement effects in selective learning. *Psychon. Sci.*, 1964, **1**, 155–156.

Ramond, C. K. Performance in selective learning as a function of hunger. *J. exp. Psychol.*, 1954, **48**, 265–270.

Spear, N. E., & Pavlik, W. B. Percentage of reinforcement and reward magnitude effects in a T maze: Between and within subjects. *J. exp. Psychol.*, 1966, **71**, 521–528.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 430–444.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Terrace, H. S. Discrimination learning with and without "errors." *J. exp. Anal. Behav.*, 1963, **6**, 1–27.

Uhl, C. N. Effects of nonrewarded forced responding on acquisition and reversal of a position discrimination. *J. exp. Psychol.*, 1966, **72**, 113–119.

(Received August 9, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 574–580

STUDIES OF DISTRIBUTED PRACTICE: XXIV. DIFFERENTIATION AND PROACTIVE INHIBITION[1]

BENTON J. UNDERWOOD AND BRUCE R. EKSTRAND

Northwestern University

This study was concerned with the fact that for the A-B, A-C paradigm, widely distributed practice (DP) on A-B will markedly reduce proactive inhibition (PI). It was assumed that this was due to increased differentiation between the 2 lists. To decrease differentiation while still maintaining DP, in some conditions pairs occurring in A-B were repeated in A-C. These operations were also carried out with the A-B, C-D paradigm, and a control was used to assess amount of PI. DP was again shown to decrease PI for the A-C paradigm in 24-hr. recall; statistically, the level of recall was equivalent to the control. No influence of DP was observed for the C-D paradigm. Repeated pairs increased PI for the nonrepeated pairs in A-C, having no influence on C-D.

In a recent study (Underwood & Ekstrand, 1966), employing the A-B, A-C paradigm, it was shown that distributing the learning of A-B over 4 days markedly reduced the proactive inhibition (PI) in the 24-hr. recall of A-C. The major concern of the present report is an evaluation of one mechanism which may be responsible for the reduction in PI. In the previous study the reduction in PI was essentially the same with 12 distributed practice (DP) trials on A-B as with 80 trials. Furthermore, 80 massed practice (MP) trials on A-B resulted in heavy PI—far more than that resulting from 12-DP trials on A-B. Therefore, it does not seem that the cause of the reduction in PI resulting from DP can be attributed to some process associated with the degree of A-B learning as such. As an alternative, the hypothesis is tested that the reduction in PI produced by DP on an interfering list results from enhanced differentiation which in turn reduces interference at recall.

Differentiation as an explanatory concept has a spotty history in forgetting theory. In recent years, however, the evidence has been mounting that such a notion may serve a useful purpose. Furthermore, there is now direct evidence that confusion (lack of differentiation) between two lists will change as a function of certain variables (Winograd & Smith, 1966). In the usual PI experiment, *S*, at the time of recall, is instructed to produce responses from the second list learned. In the A-B, A-C paradigm the stimulus words cannot provide a means of distinguishing between the two lists. Therefore, other cues must serve this function. When the learning of the AB list is distributed over 4 days, as was true in the earlier experiment, it is possible that cues associated with the temporal dispersion of the learning may allow *S* to differentiate between the two lists. That no unlearning of AB occurred suggested that differentiation must have been high at the

---

[1] This work was supported by Contract Nonr-1228 (15), Project NR 154-057, between Northwestern University and the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

time of A-C learning. If DP on A-B reduces PI because it increases differentiation between the two lists, the reduction should disappear if differentiation can be destroyed while still maintaining DP on A-B. There are several ways this might be accomplished. For example, giving DP on both A-B and A-C might reduce differentiation (Houston & Reynolds, 1965). For the present study, another technique was used. If some of the pairs appearing in A-B are maintained intact in A-C, it would seem that *S* could not differentiate the two lists on the basis of the temporal dispersion of the learning of the pairs constituting the A-B list. Therefore, conditions were arranged so that zero, three, or six pairs appearing in the first list were also present in the second list. The empirical question is whether or not the reduction in PI occurring when A-B is learned under DP will also be observed under the same DP schedule but with the same pairs occurring in both lists. If the reduction in PI produced by DP of the first list is due to increased differentiation, and if the use of common pairs in both lists destroys or reduces differentiation, the reduction of PI should not be present.

As a further test of the differentiation hypothesis, an A-B, C-D paradigm has been employed. With this paradigm, list differentiation is expected to be high because the stimuli in the two lists differ. Therefore, DP should not influence recall regardless of whether or not repeated pairs are used in the two lists.

METHOD

General design.—Nine conditions and nine different groups of 22 *S*s each were employed. All nine groups learned exactly the same second list by MP, the groups differing only in the nature of the first list or its conditions of presentation. The control group (C group) learned only the second list, with recall and relearning after 24 hr. Of the remaining eight groups, four learned under the C-D paradigm, four under the A-C. One group under each paradigm was given 32 MP trials on the first list, and another for each paradigm was given 32 trials, eight trials a day for 4 successive days, with the second list being learned immediately following the thirty-second trial on the fourth day. The remaining four groups also had the 4-day DP schedule on the first list. However, the second list for one group under each paradigm had three pairs in the second list which also occurred in the first, and the remaining two groups had six pairs in the second list which also occurred in the first. Since there were 12 pairs in the lists, the latter two groups were presented only six new pairs in the second list.

Two basic analyses may be anticipated. To determine the amount of reduction in PI produced by the DP schedule, the recall of the two standard (no repeated items) MP and DP groups for each paradigm may be evaluated against each other and against the recall of the C group. To determine the influence of repeated items in the second list, a 2 × 3 analysis may be performed with paradigm as one variable and zero, three, and six repeated items as the other. Thus, the two standard DP groups enter into both analyses.

Lists and procedure.—The 12-pair lists consisted of three-letter words as stimulus terms and two-syllable adjectives as response terms. From the previous study (Underwood & Ekstrand, 1966) the relative difficulty of the 12 pairs of the second list was known. When three and six items occurred in both lists, the items were chosen so as to represent essentially the full range of difficulty. When six items were repeated they included the three items used when three were repeated. Thus, there were six common nonrepeated items, the recall of which may be used to evaluate the effect of 0, 3, and 6 repeated items during learning.

The *S*s in the DP groups served 5 days, Monday through Friday. Eight trials were given on the first list for the first 4 days followed immediately on the fourth day by the second list which was learned to one perfect trial. Recall and relearning of the second list took place 24 hr. later. Relearning was carried to one perfect trial but with a minimum of five trials. The MP groups were given 32 trials on the first list on Thursday followed immediately by the second list, with recall of the second list 24 hr. later. All groups (except the C group) were given a free recall test (MMFR) following relearning. The stimulus terms were presented on a sheet of paper with *S* asked to supply the response terms for both lists and to identify the list they were in. No time limit was imposed.

The lists were learned and relearned by the anticipation method with a 2:2-sec. rate and a 4-sec. intertrial interval. All *S*s had had previous experience in verbal-learning experiments. The *S*s were assigned to conditions on a schedule sheet consisting of 22 blocks of the nine randomized conditions. The *S*s in the MP conditions, arriving on Monday by appointment, were asked to return on Thursday for the experiment.

RESULTS

First-list learning.—The mean number of trials to reach a criterion of one perfect trial on the first list varied from 7.82 to 12.45. However, the *F* (1.49) was far from significant. For both paradigms the scores were a little higher for the DP groups than for the MP groups, a result which might be expected since *S*s in the DP groups who failed to reach a criterion of one perfect within the first eight trials would be penalized because of forgetting between the first two sessions.

A determination was made of the number of correct anticipations given out of the total possible for the 32 trials on the first list. The percentages varied from 84.82 to 90.04. Thus, *S*s maintained a high rate of correct anticipation across the 32 trials.

Second-list learning.—The mean numbers of trials to learn the second list are plotted in Fig. 1. The performance of the C group is shown by the dotted horizontal line, and that of the two MP groups by filled circles. A statistical analysis of the six means connected by the lines showed that, as might be expected, performance improved as the number of repeated pairs increased, *F* (2, 126) = 7.37, $p < .01$. Performance under the C-D paradigm was significantly superior to that under the A-C paradigm, *F* (1, 126) = 23.02, $p < .01$, indicating negative transfer in the latter. Furthermore, it can be seen that performance under the C-D paradigm with no repeated pairs was superior to that shown by the C group, suggesting a considerable positive effect due to nonassociative factors (warm-up, learning-to-learn) as a result of first-list learning in the C-D paradigm. Although the MP groups are numerically superior to the DP groups for the standard paradigms (0 repeated pairs), the effect is far from significant statistically ($F = 1.61$). The relationships noted in Fig. 1 were essentially the same as those found when performance on the first three trials of the second list was used as the response measure.

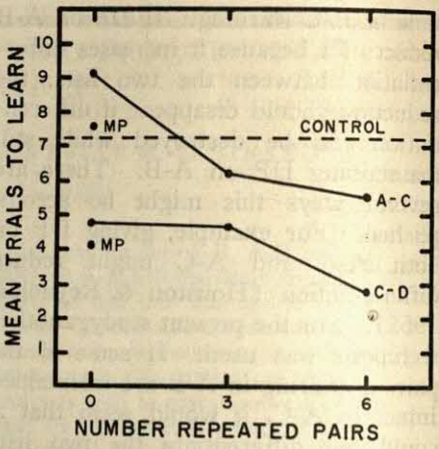


FIG. 1. Mean number of trials to learn the second list as a function of paradigm and number of repeated pairs.

Some note should be taken of the performance on the repeated pairs. For the C-D paradigm, with both three and six repeated pairs, performance was essentially perfect; over 99% of the opportunities for being correct were correct. For the A-C paradigm, for both the three and the six conditions, repeated pairs were correct on 94% of the opportunities. The same pairs were scored for this paradigm in the condition in which no items were repeated and both the three and the six items showed 75% correct responses. Thus, while there is some disturbance on the repeated pairs in the A-C paradigm, the performance on them is much better than if they had not occurred in the first list.

Recall: Standard groups.—The recall of the groups in which no items were repeated will be examined first. A plot of the recall scores is shown in Fig. 2, with MP-DP as the base line, and the prominent features of this graph will be pointed out first. As in the previous study (Underwood & Ekstrand, 1966), DP on the first list of the A-C paradigm markedly reduces PI. However, the reduction does not appear to be complete—the recall of the C group is superior to that of the DP group. Whether the first list is learned under MP or under DP has little influence on PI for the C-D paradigm, but again, the recall for both groups is less than for the C group.

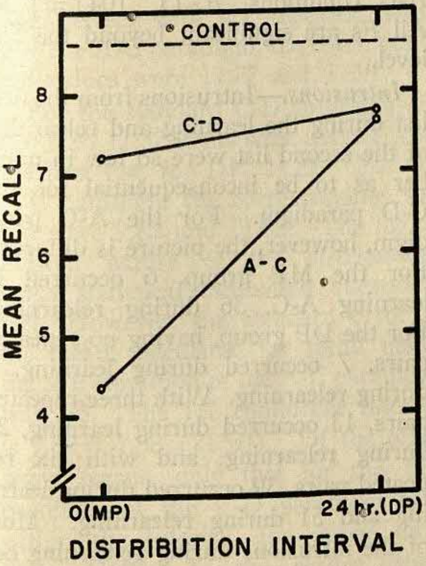


FIG. 2. Recall as a function of paradigm and distributed practice.

A statistical analysis was first performed on the four groups for the two paradigms. The effect of Paradigm, *F* (1, 84) = 8.96, MP-DP, *F* (1, 84) = 14.61, and the interaction, *F* (1, 84) = 7.89, were all significant beyond the 1% significance level. The scores for the C group were then pooled with those for the other four distributions and the error estimate used to make Duncan range tests. The outcome showed that the recall of the MP group for the A-C paradigm differed from the other four groups but the remaining four groups did not differ among themselves. Statistically, therefore, it appears that PI is not significant for the C-D paradigm nor for the A-C paradigm when DP is given on the first list.

It was shown above that learning the second list under the C-D paradigm occurred more rapidly than under the A-C paradigm. It is possible, therefore, that the level of learning of the second list may have differed somewhat for the various conditions. To adjust for such differences, a multiple-entry probability analysis was carried out and loss scores determined. The above statistical analyses were repeated using the loss scores with the outcome being that no conclusion changed except that the MP recall under the C-D paradigm was significantly less than that of the C group.

Recall: Effect of repeated pairs.—Figure 3 shows the recall for the two paradigms when zero, three, or six pairs from the first list were retained in the second. The recall scores involve only the six nonrepeated (new) pairs. For comparative purposes, the recall for these six pairs has been determined for the C group and for the MP condition of the A-C paradigm. For these six items it may be noted that the C group, all three conditions for

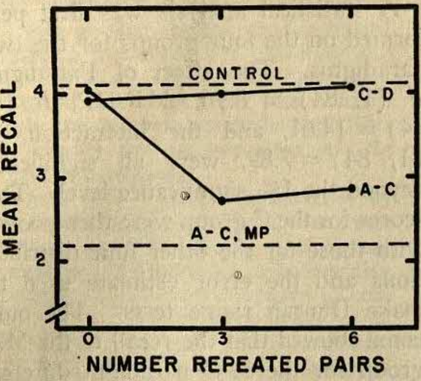


FIG. 3. Recall as a function of paradigm and number of repeated pairs.

the C-D paradigm, and the DP condition with the A-C paradigm having no repeated pairs, show virtually identical recall. This is to say, there is no evidence for PI in these conditions for these pairs. However, for the A-C paradigm, when three or six items are repeated there is heavy PI for the nonrepeated items. Roughly, the PI is two-thirds of the amount shown when the first list of the A-C paradigm is learned by MP. A statistical evaluation of the six groups shows paradigm F (1, 126) = 11.96, $p < .01$, and the interaction between paradigm and number of repeated pairs to be significant, F (2, 126) = 3.71, $p < .05$. The main effect for repeated pairs is not significant ($F = 2.43$). With repeated pairs A-C learning was more rapid than without repeated pairs. Therefore, degree of learning of the nonrepeated pairs may not be equivalent for the different conditions. However, a probability analysis performed only on the six nonrepeated items does not change the basic conclusion. Mean loss scores were 1.36, 3.00, and 2.50, for zero, three, and six A-C conditions, respectively. It is concluded that the use of repeated pairs in the C-D paradigm has had no appreciable effect on the recall of the nonrepeated pairs whereas repetition in the A-C paradigm produces a marked decrement.

For the C-D paradigm, recall of the repeated items was essentially perfect. All *S*s correctly recalled the three items when only three items were repeated. When six items were repeated, 21 of the *S*s recalled all six, the other *S* missed one. For the A-C paradigm, the mean recall of the three items was 2.55 when three were repeated, and 5.59 when six were repeated.

The relearning was so rapid, particularly in the repeated-pair conditions, that statistical analysis did not seem justified because of the nature of the distributions. However, an analysis was made of performance on the second relearning trial for the standard conditions. The findings of the recall trial (Fig. 2) were still evident. Performance was better under the C-D paradigm than under the A-C, F (1, 104) = 7.38; for A-C, MP was inferior to DP, F (1, 104) = 4.74, and the C Group's performance was better than the combined performance of the two A-C conditions, F (1, 104) = 9.20. All Fs are significant beyond the 5% level.

Intrusions.—Intrusions from the first list during the learning and relearning of the second list were so few in number as to be inconsequential for the C-D paradigm. For the A-C paradigm, however, the picture is different. For the MP group, 6 occurred in learning A-C, 36 during relearning. For the DP group, having no repeated pairs, 7 occurred during learning, 7 during relearning. With three repeated pairs, 13 occurred during learning, 27 during relearning, and with six repeated pairs, 39 occurred during learning and 51 during relearning. Most of the intrusions during relearning occurred on the recall trial. These frequencies for the conditions with repeated pairs are particularly impressive when it is recognized that with six repeated pairs not a single intrusion occurred during learning or relearning for the repeated pairs. Hence, there were only six pairs for which intrusions occurred. Of course, an intrusion occurring for a repeated item must necessarily be incorrectly paired. With three repeated items, one intrusion occurred during learning for a repeated pair, and one at recall. For this group there were, then, effectively nine pairs for which intrusions might occur.

MMFR.—Free recall of the first list may be used to assess unlearning of the first list produced by the learning and relearning of the second. Lenient or stringent scoring systems yielded essentially the same results; the data given here are the mean number of responses recalled and correctly paired. In neither paradigm is there evidence for unlearning when items are repeated in the two lists. For the C-D paradigm, 11.82 responses were correct with three repeated items and recall was perfect with six repeated items. For the A-C paradigm, the corresponding values were 11.77 and 11.91.

For the standard C-D paradigm under MP, 11.32 items were correct from the first list, for DP, 11.82. For the A-C paradigm the corresponding values were 9.82 and 11.68. The highly skewed nature of the distributions make statistical evaluation questionable. However, the standard error of the mean, calculated for recall under the MP condition of the A-C paradigm, was .56. The 1% confidence interval is 9.82 ± 1.58 (21 *df*). Of the other seven means, only that for the MP condition of the C-D paradigm falls within the interval, the upper limit being 11.40. In the earlier study (Underwood & Ekstrand, 1966) it was concluded that all MP groups showed unlearning for the A-C paradigm. The recall for the present MP group is less than for the comparable group in the previous study; it seems highly probable that unlearning did occur for this condition and that little if any occurred for any of the others.

DISCUSSION

The basic question of the present experiment concerned the role of differentiation in producing the DP effect (reduction of PI) in the A-C paradigm. It was assumed that the temporal dispersion of the learning over days resulted in high differentiation between the two lists, and that the repetition of pairs of items would reduce or eliminate this differentiation. The percentage recall for the six nonrepeated items when three pairs were repeated was 46%, and 48% when six items were repeated. For these same items for the MP group the value was 37%, for DP, 67%. Therefore, of the 30% difference between MP and DP, two-thirds may be accounted for by differentiation present in DP and not in MP. That is, this follows if it is assumed that repeating pairs completely destroys the differentiation produced by DP, and that the entire difference in the recall between MP and DP is due to a difference in differentiation. These extreme assumptions, however, leave a 10% difference in recall between MP and the repeated-pair conditions unaccounted for. This is a matter that will receive attention shortly.

Does the use of repeated pairs reduce differentiation? If frequency of intrusions is used as a direct index of loss of differentiation, there is an increase in loss from zero to three to six repeated pairs in the A-C paradigm. However, assuming that all PI is due to loss of differentiation, differentiation (based on intrusion frequency) would be poorer in the condition with six repeated pairs than in the MP condition, yet recall was 10% poorer in the latter. Furthermore, intrusions were more frequent with six repeated pairs than with three, yet there was little difference in recall. It does not appear that intrusions can be used in the present experiment in a strict quantitative manner to index differentiation. It is possible that, for whatever reason, *S*s with six repeated pairs attempt more responses (as the frequency of intrusions suggest) than do those having three repeated pairs. If there is a difference in such a response-emission threshold for the two groups, more of the items given correctly at recall may have been due to chance factors in the six-repeat condition than in the three-repeat condition.

Three possibilities may be suggested for the better recall of the repeated-pair conditions over the MP condition in the A-C paradigm. First, it is possible that the use of repeated pairs does not eliminate entirely the differentiation produced by DP. Second, the use of repeated pairs to eliminate differentiation may do so completely for the DP groups but adds a positive factor in the form of intralist differentiation, i.e., repeated and nonrepeated pairs are set apart so that intralist differentiation is higher. The use of MP on the first list with repeated pairs in the second would clearly be a test of this latter possibility. The third possibility is that DP adds some unknown positive factor over and above its effect on differentiation.

It has been suggested above that loss of differentiation between lists may be largely responsible for PI. It was assumed that the reduction in PI produced by DP on the interfering list was due to increased differentiation resulting from the temporal dispersion of the learning of the potentially interfering list. The recall indicates that this temporal dispersion may in some way produce differentiating cues so that the A-C paradigm essentially becomes a C-D paradigm in terms of interference at recall. It should be clear that the description "temporal dispersion" does not identify the cues producing high differentiation. It is possible that DP as such is not the only critical factor in the present experiment; rather, a long temporal interval (e.g., 4 days) between the onset of learning of the two lists may provide some appropriate differentiating cues to reduce PI, although available evidence (Keppel, 1964) does not indicate that this could account for all the DP effect observed here. Peterson and Gentile (1965) have shown that the longer the interval between presentation of items in short-term memory the less the PI; perhaps the present operations, using DP, are simply a variant on the temporal-separation variable. If so, differentiation may result from different tasks being associated with different "time tags" (Yntema & Trask, 1963).

REFERENCES

Houston, J. P., & Reynolds, J. H. First-list retention as a function of list differentiation and second-list massed and distributed practice. *J. exp. Psychol.*, 1965, **69**, 387–392.

Keppel, G. Facilitation in short- and long-term retention of paired associates following distributed practice in learning. *J. verbal Learn. verbal Behav.*, 1964, **3**, 91–111.

Peterson, L. R., & Gentile, A. Proactive interference as a function of time between tests. *J. exp. Psychol.*, 1965, **70**, 473–478.

Underwood, B. J., & Ekstrand, B. R. An analysis of some shortcomings in the interference theory of forgetting. *Psychol. Rev.*, 1966, **6**, 540–549.

Winograd, E., & Smith, W. S. List differentiation with varied trials on both lists. *Science*, 1966, **152**, 1101–1102.

Yntema, D. B., & Trask, F. P. Recall as a search process. *J. verbal Learn. verbal Behav.*, 1963, **2**, 65–74.

(Received August 11, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 581–585

ORDER OF PRESENTATION, CS INTENSITY, AND RESPONSE LATENCY[1]

G. ROBERT GRICE, JOHN J. HUNTER,[2] DAVID L. KOHFELD,[3] AND LARAINE MASTERS

University of Illinois

5 orders of presentation of 50- and 100-db. auditory CSs were used in a study of eyelid conditioning. The orders were random, single alternation, double alternation, and alternate blocks of 5 and 10. A 2nd single alternation group was informed of the order in advance. Regularity, instructions concerning regularity, and block length, all failed to affect the size of the intensity effect. Latency distributions are also presented. Responses to the loud tone reach their peak frequency earlier in the interval than do those to the soft. The possibility that latency mechanisms account for intensity effects in conditioning is suggested.

When Grice and Hunter (1964) reported substantially greater CS intensity effects in eyelid conditioning with within-*S* variation of intensity than with groups receiving single intensities, they proposed a number of experimental approaches directed toward further elucidation of this difference. Among these were, "the length of the trial blocks in which a particular value is consecutively presented," and "the degree of predictability in the stimulus presentation sequence." Full rationale for these suggestions was not presented. However, the proposal about block length concerned the authors' hypothesis that differences in response strength depend upon departure of the CS from adaptation level (AL). It seemed likely that if a stimulus of a particular intensity were presented a number of times consecutively, AL would move toward that level. This leads to the prediction of smaller intensity effects as run length increases.

The proposal concerning predictability was based upon no particularly sophisticated theory. The thinking was that an unpredictable, strong stimulus might have more drive or startle value than a predictable stimulus of the same intensity. Furthermore, it was thought that an unpredictable weak stimulus might have some of the characteristics of an anticlimax. It also seemed that, in the case of long run length, *S* might often be quite confident of what stimulus was coming next, and that this would strengthen any effects of mere regularity of stimulus order. Thus, both of these lines of reasoning led to the prediction of decreased intensity effects when long runs of the same stimulus are presented.

The present experiment is an attempt to investigate these two variables in eyelid conditioning with two intensities of an auditory CS. All trials were reinforced by an air-puff UCS. One irregular (unpredictable) and four regular (predictable) orders of CS presentation were used. The regular orders consisted simply of alternate blocks of 1, 2, 5, or 10 presentations of each stimulus.

---

[1] This investigation was supported by Public Health Service Research Grant MH 08033 from the National Institute of Mental Health.

[2] Now at San Francisco State College.

[3] Now at United States Army Medical Research Laboratory, Fort Knox, Kentucky.

METHOD

Apparatus and procedure.—The apparatus consisted of conventional eyelid conditioning equipment previously described (Grice & Davis, 1960). The CSs were 1,000-cps tones of 50- and 100-db. SPL, and overlapped the UCS. They were switched by an electronic switch with a rise time of 30 msec. The UCS was a 1.0-psi puff of air, 40 msec. in duration, delivered to the cornea and was presented on every trial. The interstimulus interval was 310 msec. A warning buzzer was sounded 2, 3, or 4 sec. prior to each trial, to which *S* was instructed to blink once and look at a fixation point. Conventional neutral instructions were used. The intertrial interval was irregularly 15, 20, or 25 sec.

There were 100 conditioning trials, with each CS presented on 50 trials. Five basic orders were used. One was irregular (I) with each CS occurring 10 times in each block of 20 trials. A second order was single alternation (A1) and a third was double alternation (A2). The remaining orders were alternate blocks of 5 (A5) and 10 (A10) presentations of each CS. Half of the *S*s with each order were run with the positions of the loud and soft tones interchanged to counterbalance position in the conditioning sequence. An additional group was run with the A1 order, but *S*s were informed in advance that the loud and soft tones would occur on alternate trials (A1-I).

Blinks resulting in a deflection of 1 mm. or more in the interval 150–332 msec. were scored as CRs. The *S*s for whom one-half or more of the blinks were of voluntary form (Spence & Ross, 1959) were eliminated. Also *S*s showing excessive adaptation of the UCR were eliminated. The criterion for this was five successive trials on which the UCR fell to one-third or less of its initial amplitude.

Subjects.—The *S*s were 144 women from an introductory course in psychology. There were 24 in each experimental group. Fifteen additional *S*s were eliminated because of voluntary responses, and 18 because of adaptation of the UCR.

RESULTS AND DISCUSSION

Frequency measures.—Data are presented for the last 60 trials. There were 30 presentations of each CS in this block. Mean numbers of CRs to the 50- and 100-db. tones and the differences indicating the size of the intensity effect are presented in Table 1. The initial analysis was performed omitting Group A1-I. Consistent with other within-*S* experiments (Beck, 1963; Grice & Hunter, 1964), the intensity effect is large, accounting for 93% of the between-cell variance, and is statistically significant, F (1, 115) = 290.87, $p < .001$. Furthermore, it is highly consistent over the experimental treatments, and certainly does not support the predictions that it should decrease with increasing regularity of order and increasing block length. Statistically not significant, F (4, 115) = .70, the Order × Intensity interaction accounts for less than 1% of the between-cell variance. The simple effect of order is not significant, F (4, 115) = .65.

Since the prediction concerning regularity of order was based on the notion of predictability, *S*s were questioned after the experiment concerning their knowledge of the order. It turned out that they were surprisingly unaware of the regularities. For example, in Group A1, 19 of the 24 *S*s were unable to report awareness of the simple alternation of loud and soft tones. For this reason, Group A1-I, in which *S*s were told in advance of the alternation, was added after the original experiment. This also failed to produce the predicted result. Actually, the intensity effect for Group A1-I is slightly *larger* than for Group A1. However, when

TABLE 1

MEAN NUMBERS OF CRs AND LOUD–SOFT DIFFERENCES

	Group					
	I	A1	A1-I	A2	A5	A10
Loud	15.0	13.1	15.8	15.9	16.3	16.0
Soft	7.9	6.4	8.0	8.3	8.7	7.2
L – S	7.1	6.7	7.8	7.6	7.6	8.8

Note.—Error mean squares appropriate to estimating the variability in this table are as follows: between *S*—87.6, *df* = 115; within-*S*—11.5, *df* = 115. The intensity effect and the interaction are within-*S* effects.

these two groups are compared in a separate analysis, the Instructions × Intensity interaction is not significant, $F\ (1, 46) = .40$.

What this experiment seems to show is that predictability of order does not affect the size of the within-*S* intensity effect. However, prediction and predictability are not the same thing, and it now appears that the hypothesis should have been stated in terms of actual prediction or expectancy. While the results of Group A1-I are suggestive that this also may be ineffective, they cannot be regarded as conclusive. To insure that *S* have current knowledge concerning the next stimulus, it apparently will be necessary to inform him explicitly shortly before each trial.

The prediction concerning the possible effect of block length was based on the notion that AL might shift toward the current stimulus value during the block and thus reduce the magnitude of the intensity effect. The possibility was considered that this effect might be too subtle to affect substantially the total number of CRs. Consequently, a separate analysis was made with the group with the longest block length, A10. For the last three blocks with the loud CS and the last three with the soft CS, the total CRs to each were summed, trial by trial, over the 10 trials of the three blocks combined. There was considerable variability in these trial by trial data, but it appeared possible that there was a very slight trend toward a decreasing loud-soft difference. In order to test for significance, blocks were divided into first and second halves of 5 trials each. The difference in mean numbers of CRs to loud and soft in the first half of the blocks was 4.79, and 4.07 in the second half. This difference between differences is small, but in the predicted direction. However, it is not significant as indicated by the Halves × Intensity interaction, $F\ (1, 23) = 2.57$, $p > .10$. The within-*S* intensity effect is quite stable. If it can be reduced by repeated exposure to one stimulus, it appears that it must be in blocks of substantially greater length than 10 trials.

Latency measures.—There are no data in the literature on latency of eyelid CRs as a function of CS intensity other than the categorization into long and short latency responses presented by Walker (1960) and Beck (1963). Such information is of considerable interest in view of the well-known relationships between reaction time and stimulus intensity (Woodworth & Schlosberg, 1954). Because of the high degree of homogeneity among the conditions of the present experiment, a substantial body of latency data is available for examination. Latency of all blinks occurring on Trials 41–100 were read to the nearest 8 msec., and were combined for the five groups of the original experiment. (There were no systematic differences in the forms or locations of the distributions for the groups.) Both absolute and relative frequency distributions of responses to the loud and soft tones are presented in Fig. 1. First with respect to the bimodality of these distributions, it is not unusual to observe this in eyelid data plotted in this way. In the present instance at least, the bimodality is an individual difference phenomenon rather than indicating a significant characteristic of the individual *S*'s distribution. For example, the early mode may be removed by eliminating about a quarter of the *S*s whose primary mode is in the earlier interval. This does nothing materially to change the picture, other than to make the distributions smooth backward extrapolations of the latter portions of those in Fig. 1.

There is no important difference in

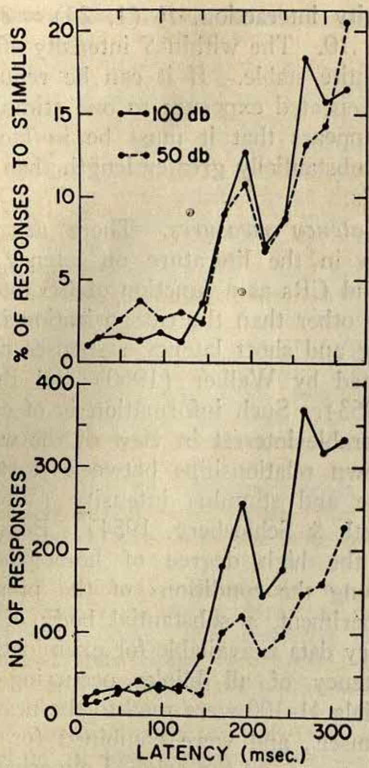


FIG. 1. Absolute and relative distributions of latency of response to the loud and soft tones.

the central tendencies of the distributions for the loud and soft tones. Computed from the grouped data for responses in the CR range only, the median for the loud tone is 258 msec. and is 261 msec. for the soft tone. The means are 247 and 252 msec. for the loud and soft tones, respectively. An inspection of the forms of the distributions is somewhat more suggestive. The distribution of responses to the loud tone reaches its peak at about 260 msec., while that for the soft tone continues to rise until the very end of the interstimulus interval. This suggests that there may be a more sizable effect of intensity on latency which is obscured by the end of the interstimulus interval. It also suggests the very interesting possible interpretation that the effect of stimulus intensity upon the frequency of short latency CRs may be deducible from latency effects and the duration of the interstimulus interval. Of course, extrapolation of the present functions would be merely speculative. While the interval used in this experiment represents the range in which Walker (1960) found the maximum effect of intensity, it is now apparent that future investigations of this relationship will require the use of longer intervals.

Of particular interest to the authors, is the possible relation of data such as these to decision models of latency phenomena such as the one proposed by McGill (1963). Briefly, such a model may state that a sensory input consists of temporal sequence of impulses, the rate of which is dependent upon stimulus intensity. When the number of impulses reaches some fixed value, *S*s detection criterion, *S* will respond. McGill has used the model to account for the negative relation between reaction time and stimulus intensity, but his primary concern has been with the stochastic properties of the model and its distributional implications. It now appears that this is a promising approach to intensity and interstimulus-interval phenomena in conditioning. Furthermore, if one assumes that *S*'s' criterion can be experimentally manipulated, the model will, in principle, account for the Grice and Hunter (1964) effect, and for the effect of preadaptation reported by Murray and Kohfeld (1965). These implications are to be discussed more fully in another place.

REFERENCES

BECK, S. B. Eyelid conditioning as a function of CS intensity, UCS intensity, and Manifest Anxiety scale score. *J. exp. Psychol.*, 1963, **66**, 429–438.

GRICE, G. R., & DAVIS, J. D. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *J. exp. Psychol.*, 1960, **59**, 391–395.

GRICE, G. R., & HUNTER, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychol. Rev.*, 1964, 71, 247–256.

MCGILL, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 309–360.

MURRAY, H. G., & KOHFELD, D. L. Role of adaptation level in stimulus intensity dynamism. *Psychol. Sci.*, 1965, **3**, 439–440.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

WALKER, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. Rev. Ed. New York: Holt 1954.

(Received August 11, 1966)

(Continued from page 561)

Image as a Mediator in One-Trial Paired-Associate Learning: B. R. Bugleski,* Edward Kidd, and John Segmen: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

Anagram Solving as a Function of Letter-Sequence Information: Roger L. Dominowski*: Department of Psychology, University of Illinois at Chicago Circle, P.O. Box 4348, Chicago, Illinois 60680.

Mediated Transfer in Reversal and Nonreversal Shift Paired-Associate Learning: Barbara W. Marquette and L. R. Goulet*: Department of Psychology, West Virginia University, Morgantown, West Virginia 26506.

Retroactive Inhibition with Different Patterns of Interpolated Lists: Judith Goggin*: Human Performance Center, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

Implicit Verbal Chaining as the Basis of Transfer in Paired-Associate Learning: Jack Richardson*: Department of Psychology, Harpur College, State University of New York at Binghamton, Binghamton, New York 13901.

Effect of "Rights" and "Wrongs" on Concept Identification: Coleman Merryman,* Barbara Kaufmann, Eric Brown, and Joel Dames: Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Associative Structure and the Temporal Characteristics of Free Recall: Howard R. Pollio,* Richard A. Kasschau, and Harry E. DeNise: Department of Psychology, University of Tennessee, Knoxville, Tennessee 37916.

Generalization and Free Recall of Similar and Opposite Words: James H. Koplin,* Danny R. Moates, and Judith Burroughs: Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

Word-Fragments as Aids to Recall: The Organization of a Word: Leonard M. Horowitz,* Margaret A. White, and Douglas W. Atwood: Department of Psychology, Stanford University, Stanford, California 94305.

Classical Conditioning of the Galvanic Skin Response to Verbal Concepts: S. Joyce Brotsky*: Department of Psychology, San Fernando Valley State College, Northridge, California 91324.

Reversal and Nonreversal Shifts among Liberian Tribal People: Michael Cole,* John Gay, Liberia Glick, and Joseph Glick: Department of Psychology, University of California, Irvine, California 92664.

Effects of Delayed Auditory Feedback on Immediate and Delayed Recall and Recognition: Charles H. D. Williams and Gerald Frincke*: Institute of Human Learning, University of California, Berkeley, 2241 College Avenue, Berkeley, California 94720.

Anxiety and Verbal Concept Learning: Ralph F. Dunn*: Department of Psychology, Loyola University, 6225 North Sheridan Road, Chicago, Illinois 60626.

Shifts in Percentage of Reinforcement Viewed as Changes in Incentive: Calvin M. Leung and Glen D. Jensen*: Department of Psychology, DePaul University, 1219 West Fullerton Avenue, Chicago, Illinois 60614.

(Continued on page 590)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 586–588

SUPPLEMENTARY REPORTS

CODING PROCESSES IN VERBAL LEARNING AS A FUNCTION OF RESPONSE PRONOUNCIABILITY[1]

WILLIAM E. FORRESTER AND NORMAN E. SPEAR
Rutgers University

The effect of rated trigram pronounciability (PR) on the extent to which Ss will use anagram coding in learning a list for free recall was investigated. Variables were coding instructions, letter-sequence restrictions on recall, and PR. All variables affected the extent and effectiveness of encoding to words. The operant level of anagram coding showed a significant inverse relationship with PR. The suggestion was made that pronunciation itself is a type of code, used whenever applicable.

Underwood and Keppel (1963) investigated the use of anagram coding in a free-learning situation. This code involves the rearranging of response letters to form a more integrated unit. Underwood and Keppel presented a list of nonword trigrams to two groups of *S*s. One group was told of the possibility for letter rearrangement to form English words, but the other group was not. Within each of these two main conditions, two subgroups were differentiated on the basis of restrictions imposed on recall: one group was required to recall letters within each item in the exact (decoded) order in which they were presented (Group 5); the other group could recall in free (encoded) letter order (Group 0). Under free letter-order conditions, instructed *S*s (who did recall 97% of their responses in encoded form) performed significantly better than noninstructed *S*s. Under exact letter-order conditions, however, noninstructed *S*s performed significantly better than their instructed counterparts. The latter result was attributed to interference effects at decoding. Finally, the group not instructed on coding but allowed to recall trigram letters in any order did, nevertheless, recall 56% of their responses in the form of words. This latter measure may be termed the operant level for coding by this method.

If *S*'s reason for using such coding devices is to increase response integration, then the higher the initial level of meaningfulness of list items, the less anagram coding should occur. The effects found by Underwood and Keppel and the operant level of anagram coding should thus be diminished at higher levels of meaningfulness. To test this hypothesis, the general design and procedures were similar to those of Underwood and Keppel (1963), but trigram meaningfulness was also varied between two lists. The specific index of meaningfulness manipulated was that of rated pronounciability (PR) as determined by a table of PR of 66 trigrams developed at Rutgers (Forrester, 1966). Ratings were on a 9-point scale with 9 representing extremely low PR.

Method.—Eighty introductory psychology students, naive to verbal-learning experiments, were assigned at random to each of the eight independent groups ($n = 10$) dictated by a $2 \times 2 \times 2$ factorial design. The factors involved were PR—High (H) vs. Low (L), instructions to induce coding (I) vs. No such instructions (N), and Letter-order required at recall—Exact (E) vs. Free (F).

The specific items used in List H were: IRM, TUC, ELD, SNU, NEP, UMB, and DAP (Mean $PR = 2.54$). Corresponding items in List L were: RMI, UTC, LDE, NSU, ENP, UBM, and PDA (Mean $PR = 7.04$). To equate the lists for intralist similarity and codeability, each item in one list corresponded to an item in the other list in that the same letters were used, but the order differed. Ease of encoding was further controlled by equating the lists for number of coding rules (e.g., "Put the last letter between first two," etc.). All items could be transformed into English words by rearranging letters. In terms of generated values (cf. Underwood & Schulz, 1960), List H ($\bar{x} = 24.86$) was slightly

---

[1] This paper is based on a thesis submitted by the first author to the Graduate School of Rutgers University in partial fulfillment of the MS degree. The research was supported in part by grants to the second author from the National Institute of Mental Health (Grant MH10194) and the Rutgers University Research Council.

lower than List L ($\bar{x} = 26.29$)—a nonsignificant difference.

Items were presented at a 3-sec. rate on a Stowe memory drum. All *S*s had five trials with a 40-sec. interval for unpaced written recall after each. A different random order of items was used on each trial.

Results and discussion.—The data, consisting of responses summed across trials, are displayed in Table 1 where they are categorized according to letter order of recalled items. Only those responses in which an item's three correct letters were given by *S* are included in the table.

The first analysis was on total correct responses (i.e., exact letter order in Cond. E vs. any order in Cond. F). As should be expected from the differential criteria imposed on these two conditions, Cond. F was superior to Cond. E. An analysis of variance revealed, however, that there was a significant Letter Order × PR interaction, F (1, 72) $= 5.04$, $p < .05$. Separate analyses of variance at each level of PR showed the nature of the interaction to be such that Cond. F was significantly superior to E only under low PR conditions, F (1, 36) $= 12.99$, $p < .001$. The instructional variable had no reliable ($p > .05$) effect on total correct responses.

In order to determine if the differences found under the total-correct measure were due to the fact that different criteria were used in scoring correct responses, additional analyses were performed. The first included only those responses recalled in exact letter sequences, and the second included only those responses recalled in nonexact letter sequences. The major results were that High PR groups were superior to Low PR groups in recalling exact letter-order responses, F (1, 72) $= 22.99$, $p < .001$, while low PR groups tended to recall more nonexact letter-order responses, F (1, 72) $= 3.17$, $p < .10$. Use of the Newman-Keuls test (cf. Winer, 1962, pp. 80–85) showed that E groups recalled significantly more exact sequences and significantly less nonexact sequences than did the F groups. Furthermore, these differences were significantly increased when coding instructions were given to *S*s in Cond. F, but such instructions had no effect on Cond. E.

In terms of words emitted at recall, it was found that practically none (.002%) were given under Cond. E, but 47% of all responses under Cond. F were in the form of words. In the latter condition, differences between Groups I and N (67% and 25% words, respectively) were highly significant, F (1, 36) $= 12.90$, $p < .001$. The PR effect was not significant.

The results discussed above indicate that allowing *S*s to recall without letter-order restrictions did ***not*** facilitate performance when list items were of high PR. Under high PR conditions, none of the results reported by Underwood and Keppel (1963) were found. This is presumably due to the fact that high PR items were not coded to words—a presumption supported by the analyses of exact and nonexact responding. Given the Low PR list, however, Free Letter-order conditions did improve performance by increasing the amount of coding under these conditions. Instructions further increased the amount of coding in Cond. F and thereby facilitated performance—in accord with Underwood and Keppel's results.

The only factor which had a significant effect on performance under Exact Letter-order conditions in the present study was that of PR. The finding that performance in this condition was not affected by coding instructions disagrees with Underwood and Keppel (1963) who found instructions to

TABLE 1

MEAN NUMBER OF ITEMS RECALLED OVER FIVE TRIALS WITH CLASSIFICATION ACCORDING TO LETTER ORDER WITHIN A GIVEN ITEM

	Free Letter Order				Exact Letter Order			
	Instructed		Not Instructed		Instructed		Not Instructed	
	High PR	Low PR	High PR	Low PR	High PR	Low PR	High PR	Low PR
Exact Order	11.00	3.90	22.40	13.60	26.20	15.90	25.70	18.20
Words	18.00	19.50	4.00	8.00	0.10	0.00	0.10	0.00
Other Order	0.10	3.60	0.10	1.70	0.50	3.30	0.10	0.40
Totals	29.10	27.00	26.50	23.30	26.80	19.20	25.90	18.60

have a significant inhibitory effect under exact letter-order procedures. Reasons for this discrepancy are unclear but may be connected to factors mentioned below in the discussion of operant level.

Although PR had no overall effect on the recall of words under Cond. F, it did affect the operant level (i.e., relative frequency) of such coding in the noninstructed subgroups of this condition. By using the formula $\frac{\text{words}}{\text{total correct}}$ to derive this term, the operant level of Group N-F in the present study was found to be approximately 24%. Considering the PR factor, the operant levels for H-N-F and L-N-F were 15% and 34%, respectively. Thus even under List L, the operant level fell below the 56% found by Underwood and Keppel. One factor which may account for this difference is the fact that each item used by Underwood and Keppel could be transformed into *two* English words. It may also be that items used by those authors are even lower in PR than the ones used in the present study.

In a final statistical analysis, the 14 trigrams used in the present study were ranked in terms of the PR of each item and the operant level of coding for each trigram. A rank-order correlation between the two measures showed a significant inverse relationship ($rho = -.75$, $p < .05$). This result, together with the finding that the overall frequency of coded responses tended ($p < .10$) to be inversely related to trigram pronounciability, lends credence to the hypothesis that pronunciation itself is a form of coding. The present results thus strengthen the notion of "sound" coding which Underwood and Erlebacher (1965, Exp. VI) invoked in interpreting the results of their transfer study.

REFERENCES

Forrester, W. E. Coding processes in verbal learning and the role of pronounciability of response items. Unpublished masters thesis, Rutgers University, 1966.

Underwood, B. J., & Erlebacher, A. H. Studies of coding processes in verbal learning. *Psychol. Monogr.*, 1965, **79**(13, Whole No. 606).

Underwood, B. J., & Keppel, G. Coding processes in verbal learning. *J. verbal Learn. verbal Behav.*, 1963, **1**, 250–257.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 16, 1966)

Journal of Experimental Psychology
1967, Vol. 74, No. 4, 588–590

EXPERIMENTALLY INDUCED RESPONSE BIASES AS A FUNCTION OF POSITIVE AND NEGATIVE WORDING

MICHEL HERSEN[1]

Psychotherapy Associates of Fairfield, Fairfield, Connecticut

The present investigation was conducted to determine whether the experimental induction of response biases, demonstrated to occur in other studies, is a function of positive and negative wording. The experiment was designed to examine whether Ss, consistent with their training series history, respond according to these specific aspects of the items in testing. The results indicate that the direction of wording in testing is a factor that either facilitates conditioning or acts as a suppressor variable, irrespective of S's ability to specify the experimental arrangements.

Recently Bugelski and Hersen (1966), Hersen (1967a, 1967b) have demonstrated that positive and negative response biases can be experimentally induced (*S*s "conditioned" to respond true or false to "ambiguous" statements in training and tested with similar "ambiguous" items in extinction), and further that such response biases can generalize from one concept topic to another. Although the results obtained from these experiments are fairly conclusive, there is a question as to the generality of the findings, particularly regarding the effect of specific aspects of the individual items. The present study was designed to determine whether *S*, consistent with his reinforcement history in training, responds during testing to ambiguous items in terms of positive and negative wording.

---

[1] Data for this study were collected while the author was at the Psychology Department of the State University of New York at Buffalo.

Method.—Men and women from introductory psychology classes from the State University of New York at Buffalo participated as *S*s in a group setting. There were 24 *S*s in the Positive Experimental group, 25 *S*s in the Positive Control group, 27 *S*s in the Negative Experimental group, and 22 *S*s in the Negative Control group.

Materials, instructions, setting, and procedure were essentially the same as used for the Experimentals and Control Group II, described in the Bugelski and Herson (1966) study. However, the statements on "Old Age" were slightly modified in terms of positive and negative wording. In the Positive Experimental group (16 ambiguous, 16 true, 16 false statements in random order) all ambiguous items, positively worded, were verbally reinforced as true by *E*. The true and false statements were also reinforced accordingly by *E*. The ambiguous items comprising the test series (16 ambiguous, 16 true, 16 false in random order) were worded positively. None of the test series items were reinforced. The training series for the Positive Controls included 24 true and 24 false statements reinforced as such by *E*. The test series was the same as used for the Positive Experimentals. For the Negative Experimentals and Negative Controls the same training series as described above was employed. However, the test series for these two groups included ambiguous items worded *negatively*, thus providing the basis for comparison.

Following the test series *S*s were administered a questionnaire to determine their knowledge of the experimental arrangements (Verbalizers or Nonverbalizers).

Results and discussion.—The major findings are summarized in Table 1.[2] An examination of the table indicates that Positive Verbalizers and Negative Verbalizers performed better than their respective Nonverbalizer counterparts. However, a comparison of the Positive Nonverbalizer group with the Positive Control yields a *t* of 2.41 ($p < .05$), indicating that in this instance conditioning took effect even though *S*s were unable to report the response-reinforcement contingency.

With respect to the main hypothesis being tested the last three comparisons presented in Table 1 are most pertinent. The difference between the Positive Verbalizers and Negative Verbalizers shows that, although verbal conditioning conducted with positively worded statements transfers reliably to negatively worded statements, the negative wording in the test series has a suppressing effect on performance. This was also confirmed by a comparison of Positive Nonverbalizers and Negative Nonverbalizers ($t = 2.93$, $p < .01$) where a similar suppressor variable seems to be operating, in spite of the fact that these *S*s were unable to verbalize the correct contingency. It could be argued that the differences are due only to a general "yeasaying" tendency towards positively worded statements and a "naysaying" tendency towards negatively worded statements. However, this is contraindicated by the nonsignificant difference between the two control groups where no ambiguous items appeared in training.

TABLE 1

Comparison of Mean Number of "True" Responses to Ambiguous Items in the Test Series Made by the Experimentals and Controls

Group	*N*	*M*	*SD*	*t*
Positive Verbs vs. Positive Control	10 25	12.50 7 56	1.57 2.32	6.59***
Positive Nonverbs vs. Positive Control	14 25	9.39 7.56	2.45 2.32	2.41*
Positive Verbs vs. Positive Nonverbs	10 14	12.50 9.39	1.57 2.45	3.11**
Negative Verbs vs. Negative Control	9 27	10.00 6.93	2.15 2.02	3.71***
Negative Nonverbs vs. Negative Control	13 27	6.64 6.93	2.02 2.02	.26
Negative Verbs vs. Negative Nonverbs	9 13	10.00 6.64	2.15 2.02	3.59**
Positive Verbs vs. Negative Verbs	10 9	12.50 10.00	1.57 2.15	2.63*
Positive Nonverbs vs. Negative Nonverbs	14 13	9.39 6.64	2.45 2.02	2.93**
Positive Control vs. Negative Control	25 27	7.56 6.93	2.32 2.02	.95

* $p < .05$.
** $p < .01$.
*** $p < .001$.

In addition, examination of performance on true and false statements for all groups, did not show the effects of conditioning. Finally, comparison of training series data for the two experimentals failed to yield significant differences, hence no indication of sampling bias or *E* bias.

The significant differences between the Positive Verbalizers vs. the Negative Verbalizers and Positive Nonverbalizers vs. the Negative Nonverbalizers support a cognitive interpretation of the data. That is, *S* in responding to the ambiguous items takes cognizance of particular aspects of each item and responds accordingly. The response is not merely one of true to ambiguous, but it is a response of true only when the item is worded positively in testing as it was in training. Conversely, for the Negative Experimentals in testing, *S*, consistent with his reinforcement history, responds false (double negative) to maintain the positive response bias. Thus, the Negative Experimentals obtained fewer "true" responses in testing which shows that a suppressor variable (negative wording) was in operation. In essence, these data show that *S*s do respond to very specific aspects of the statements.

---

[2] Data for males and females were combined as no sex differences were found.

Finally, it is important to note that differences occurred even when some *S*s (nonverbalizers) were unable to identify the experimental arrangements. This, however, does not detract from the results since verbalization of answer contingencies is not necessarily linearly related to performance. Furthermore, evidence accrued by Hersen (1967a) indicates that verbalization of contingencies is largely a function of questionnaire specificity.

REFERENCES

Bugelski, B. R., & Hersen, M. Conditioning acceptance or rejection of information. *J. exp. Psychol.*, 1966, **71**, 4, 619–623.
Hersen, M. Generalization of positive and negative response biases. *J. exp. Psychol.*, 1967, in press. (a)
Hersen, M. Repeated inquiry during training in a verbal conditioning paradigm. *J. gen. Psychol.*, 1967, in press. (b)

(Received July 18, 1966)

(Continued from page 585)

Choice Behavior in a Verbal Recognition Task as a Function of Induced Associative Strength: Seymour Rosenberg* and Lawrence Donner: Department of Psychology, Rutgers University, New Brunswick, New Jersey 08903.

Interactions between Items in Visual Search: Ian E. Gordon*: Department of Psychology, Queen's College, Dundee, Scotland.

Effects of UCS Adaptation upon Conditioning in Low and High Anxiety Men and Women: G. N. Hobson*: Department of Psychology, University of Victoria, Victoria, British Columbia, Canada.

Sequential Probabilities and the Performance of Serial Tasks: Don Trumbo,* Merrill Noble, and Jane Quigley: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Effects of Repetition of Voluntary Response: From Voluntary to Involuntary: In-mao Liu*: Department of Psychology, National Taiwan University, Taipei, Taiwan, Republic of China.

Visual Beat Phenomena as an Index to the Temporal Characteristics of Perception: Rathe Karrer*: Illinois State Pediatric Institute, 1640 West Roosevelt Road, Chicago, Illinois 60608.

Information Seeking when Available Information is Limited: Gordon F. Pitz*: Department of Psychology, Southern Illinois University, Carbondale, Illinois 62903.

Relationship between Static and Dynamic Stereo Acuity: S. M. Luria* and Seymour Weissman: Vision Branch, U. S. Naval Submarine Medical Center, U. S. Naval Submarine Base, New London, Groton, Connecticut 06340.

Hypothesis Behavior in a Concept-Learning Task with Probabilistic Feedback: Steven P. Rogers and Robert C. Haygood*: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Effects of Temperature and Time of Day on Time Judgments: Donald Pfaff*: Apartment 2A, 653 East 14th Street, New York, New York 10009.

*Asterisk indicates author for whom address is given.

Journal of Experimental Psychology
Monograph

Vol. 74, No. 4 • August 1967
Part 2 of 2 Parts (Whole No. 637)

EFFERENCE AND THE CONSCIOUS EXPERIENCE OF PERCEPTION[1]

LEON FESTINGER
Stanford University
HIROSHI ONO
University of Hawaii
AND
CLARKE A. BURNHAM
University of Texas
DONALD BAMBER
Stanford University

A historical review of past attempts at formulating theories in which efference plays a role in conscious perception is presented. A testable version of such a theory is formulated, and 4 experiments are presented testing implications from this theory. In all of these experiments, conditions in which Ss must learn a new afferent-efferent association are compared with Ss whose physical activity and perceptual experience are very similar but who need not learn a new association between afferent input and relevant efferent output. In all of the experiments significant change in the visual perception of curvature was obtained in the conditions in which the new associations had to be learned. Where no new associations had to be learned, significantly less visual change occurred. The results are consistent with the theoretical position that the efference and efferent readiness activated by visual input helps determine the visual perception of contour.

HISTORICAL INTRODUCTION

Around the turn of the last century there were a few people who proposed that motor output was essential to the conscious experience of perception. Münsterberg (1899), for example, elaborated the view that incoming, afferent stimulation and outgoing motor innervation were a single, continuous nerve process with no point of separation between them. The motor dis-

---

[1] The research reported in this monograph was conducted while all the authors were at Stanford University. While all of us worked together, different ones were involved in different experiments. Experiment III is a condensed version of the dissertation submitted for the PhD at Stanford University by Clarke A. Burnham. Experiment I was conducted by Festinger, Burnham, and Bamber; Ono had primary responsibility for Exp. II; Festinger and Burnham were involved in Exp. IV.

We would like to thank Lance Kirkpatrick Canon and Stanley Coren for their help in the experimental work and Douglas H. Lawrence and Charles R. Hamilton for comments, criticisms, and general assistance.

All the research was supported by grants to Leon Festinger from the National Institutes of Health, Grant No. MH 07835, and the National Science Foundation, Grant No. G-11255.

© 1967 by the American Psychological Association, Inc.

charge, he held, is necessary before any central activity corresponding to perception or consciousness takes place. Montague (1908) stated that "Perceptions are presumed to arise synchronously with the redirection in the central nervous system of afferent currents into efferent channels [p. 128]."

Ultimately, such views fell into disrepute for a variety of reasons. For one thing, these theories arose out of an attempt to understand consciousness and, with the rise of behaviorism, consciousness became less and less a proper subject for study. Thus, for example, Washburn (1916) devoted most of the introduction of her book to attacking Watson and behaviorism and to justifying that the study of consciousness is proper. But this book is one of the last attempts to state and elaborate a motor theory of consciousness.

Another reason for the decline of such theories was that they never were able to cope adequately with the facts. If the conscious experience of perception is occasioned by the motor discharge initiated by the afferent input, then one would expect all perception to be accompanied by motor movements. Münsterberg (1900) explained that we frequently do not perceive something if our attention is directed elsewhere because the appropriate motor responses are suppressed if attention is fixed on something else. He, consequently, stated that the vividness of conscious experience is a direct function of how free the motor pathway is to discharge. This view, however, seems to conflict with the fact that when a movement is well learned and occurs freely and easily, consciousness of the movement decreases. A skilled violin player, for example, is not conscious of all his movements. This led Montague (1908) to propose, contrary to Münsterberg, that consciousness is more vivid if the motor output is interfered with. Washburn (1916) attempted to reconcile all this in the statement that "consciousness accompanies a certain ratio of excitation to inhibition in a motor discharge and if the amount of excitation either sinks below a certain minimum or rises above a certain maximum, consciousness is lessened [p. 25]."

Whichever of the above hypotheses one chooses, however, one must still search for movement correlates of perception. Even if the motor discharge is interfered with somewhat, there would still be some movement. The general absence of obvious movement accompanying conscious perception led to the necessity to postulate the existence of rudimentary or tentative movements. Thus Breese (1899), in discussing the perception of speech, stated:

> The muscles of the vocal cords, throat, and respiratory organs are slightly innervated and adjusted, but the process goes no further. Sometimes, however, the enunciation is complete so far as the adjustment of the muscles of the vocal cords, throat and mouth cavities is concerned. There is a tendency to make these adjustments not only when we hear spoken words, but to make them in response to other stimuli. We are likely to utter the name of any object upon which the attention rests. . . . If, for any reason, the motor apparatus does not respond properly, there is an interruption in the conscious stream [p. 49].

The relationship between the conscious experience of perception and such motor movements, however, remained hypothetical and, to modern psychologists, implausible.

Another probable reason for the demise of these views concerning the importance of motor action for the conscious experience of perception is the unclarity concerning what explanatory power was added by the insistence that motor innervation had to exist for consciousness to exist. Up to the end of the nineteenth century there were many psychologists who accepted Helmholtz's (1962) view that there was a consciousness concerning innervation to the muscles and, as long as this view was held, motor theories of consciousness added something. James (1890), followed by Sherrington (1900), attacked this view successfully, however, and persuaded psychologists and physiologists alike that there was no conscious sense of innervation to the muscles and that afferent input via sense organs was the only input relevant to the conscious experience of perception. Thus, Münsterberg (1899) said:

> The only theory which brings in a really new factor is the theory of innervation feelings. This well-known theory claims that one special group of conscious facts, namely the feelings of effort and impulse, are not sensations and, therefore, not parallel to the sensory excitements, but are activities of consciousness and parallel to the physiological innervation of a central motor path. . . . The psychologist can show (however) that this so-called feeling of effort is merely a group of sensations like other sensations, reproduced joint and muscle sensations which precede the action. . . . If the other sensations are accompaniments of sensory excitements in the brain the feelings of impulse cannot claim an exceptional position [p. 443].

If, as became widely accepted, there was no consciousness of innervation to the muscles and all perception depended upon afferent input, then it became unclear as to the value of a theory that states that innervation to the motor system is necessary for conscious experience.

More recently, however, evidence has accumulated that, in spite of having won the argument, James (1890) and Sherrington (1900) were wrong and that the organism *does* have usable conscious information about innervation to the motor system, that is, about efferent impulses issued from the central nervous system through the motor pathways. Reviews of the issue and of the evidence may be found in Merton (1964) and in Festinger and Canon (1965). Consequently, it may be worthwhile to examine once more the possible validity of some form of theory of the conscious experience of perception which depends in whole or in part on efference activated by afferent input and to see if there are any facts which would strongly argue in the direction of some such view.

Some Observation on Perception of Limb Movement

There are some observations reported in the literature, all of them made incidentally while investigating some other problem, which seem to point in the direction of a strong effect of efference on perception. Perhaps the most interesting of these was reported by Gibson (1933) in connection with an experiment on visual adaptation to curvature. Having noted previously that if *S* wore wedge prism spectacles that made vertical straight lines appear curved, after a while *S* adapted to the curvature so that the vertical lines looked less curved, he set out to explore the mechanism of this visual adaptation. His initial hypothesis was that the adaptation occurred as a result of "conflict between vision and kinesthesis [p. 4]." That is, since the vertical lines looked curved but, if felt would feel straight, this conflict might lead to the adaptation. His *S*s, consequently, each spent about $\frac{1}{2}$ hr. wearing the prism spectacles and looking at and running their fingers along vertical edges such as a meter rule.

The observation of relevance to us here is reported in the section on procedure as follows:

> It was discovered, however, that in actual fact the kinaesthetic perception, in so far as it was consciously represented, did *not* conflict with the visual perception. When a visually curved edge such as a meter stick was felt, it was felt *as curved.* This was true as long as the hand was watched while running up and down the edge. If the eyes were closed or turned away, the edge of course felt straight, as it in reality was. This dominance of the visual over the kinesthetic perception was so complete that when subjects were instructed to make a strong effort to dissociate the two, *i.e.* to "feel it straight and see it curved," it was reported either difficult or impossible to do so [pp. 4–5].

This phenomenon reported by Gibson (1933) is clear and compelling and anyone who has a pair of prism spectacles can demonstrate it for himself. It has been tried in our laboratory, again and again. Wearing the spectacles and running one's hand up and down along, say, a door edge or door frame edge, the hand *feels* that it is moving in a curved path. It is not that one thinks the hand is moving in a curve because one sees a curve. The hand actually feels it is moving in a curve in spite of the fact that it actually is moving in a straight path. To say, as Gibson said, that visual perception dominates over kinesthetic perception does not explain the phenomenon. The question still remains as to how vision dominates proprioception so that the hand actually feels that the path of movement is curved.

If one thinks in terms of some theory in which efference affects perception, however, an explanation readily suggests itself. Let us imagine that the conscious perception of the path of movement of a limb is not the organization of informational input from the receptors in that limb, but is rather the organization of the efferent signals issued from the central nervous system *to* that limb. The arm would be felt to move in a curved path if the efferent signals issued through the motor pathways directed the arm to move in a curve. The fact that the arm and hand, because they are maintaining pressure on the straight edge, actually move in a straight line would then be irrelevant to the conscious experience of path of movement. The arm is felt to move as it has been directed to move.

From such a point of view one can understand the dominance of vision over proprioception in this instance. The dominance exists as it does because the visual input, perhaps because of years and years of prior learning, perhaps because of its greater precision, is heavily relied on to activate efferent instructions. If the visual input corresponds to a curve, then the efferent instructions activated by the input direct the arm to move in a curved path and the arm is felt to move that way. If this is true, one would expect to be able to observe some manifestation of the fact that the arm has been directed to move in a curved path and, indeed, one can observe indications of this. Typically, in this situation in which a person sees an actually straight edge as curved and runs his hand up and down along it, the wrist and hand twist somewhat in a manner consistent with directions to move in a curved path.

The view that the conscious perception of the movement of a limb is determined, or at least affected, by the efferent instructions issued from the central nervous system to that limb leads us to expect other observable phenomena. To take a very gross example, let us imagine that, using some drug, one paralyzed a person. We would expect that if this person tried to move, even though he actually did not move because of the paralysis, he would feel that he had moved. I know of no systematic data that have ever been collected in such a situation but an incidental observation by Campbell, Sanderson, and Laverty (1964) tends to support this expectation.

The authors report an experiment in which five *S*s were injected with succinylcholine chloride dihydrate. This is what the authors said about the action of the drug:

> The drug acts so as to break the connection between the motor neurones and the skeletal musculature. . . . During the period in which the drug is active the skeletal musculature is very nearly completely paralyzed. . . . (The drug) has no anesthetic effect. Enquiries made of subjects following the paralysis indicate that they are aware of what is going on around them. . . . [p. 628].

The drug produced a very traumatic experience for *S*s, largely because of the interruption of respiration. The average duration of this respiratory paralysis was about 100 sec. The incidental observation that is of interest to us here is the following:

> The subjects described their movements (during the paralysis) as part of a struggle to get away from the apparatus and to tear off the wires and electrodes. Though in fact their movements were small and poorly controlled the subjects were under the impression they had been making large movements [p. 632].

Surely, it seems difficult to imagine any basis for this phenomenon other than that their conscious experience of movement of their limbs was based on the efferent output to those limbs.

A similar instance, not involving paralysis, was reported by Merton (1964) in connection with experiments to demonstrate that joint receptors rather than muscle receptors carry information concerning the position of a limb. He stated:

> In some recent unpublished experiments, Dr. T. Davies, Mr. A. J. M. Butt and I have used the top joint of the thumb. The advantage of this joint is that the muscles that flex and extend it both lie in the forearm and have long tendons. Hence it is possible with a pneumatic tourniquet around the wrist to make the joint and the skin of the thumb anaesthetic without any effect on the muscles.
>
> This experiment succeeds in making the top joint of the thumb show just the same properties as the eye as regards movements. After an hour to an hour and a half of aschaemia the subject becomes quite insensitive to passive movements of the joint of whatever range or rapidity. Nevertheless, active movements of the joint are made with much the same accuracy as before and, indeed, with much the same angular accuracy as eye movements in the dark. *If the movement is restrained by holding the thumb the subject believes he has moved it just the same* [pp. 393–394, italics ours].

Surely, it seems difficult to imagine any basis for these observations other than that the conscious experience of movement of the limbs, in one case, or the thumb, in the other instance, was based on the efferent output to the motor system.

The Problem of a Motor Theory of Visual Perception

We have seen that there are instances in which the perception of a motor movement seems to be based on efferent output rather than on afferent input. If the discussion were to be confined to the perception of motor movements, there does not seem to be much difficulty in specifying and maintaining an "efference" theory of perception. If, however, we wish to broaden our considerations to include visual perception, we immediately encounter difficulties. Efferent instructions issued to the muscles associated with the eye do not seem to be integral in visual perception. It is difficult to imagine what specific eye movements would be relevant to the perception of brightness and color, in the first place. But even if we ignore brightness and color and think only of the visual perception of shape, pattern, and contour, there are still problems. We do not have to move our eyes along a contour in order to perceive that contour. Even if a steady point of fixation is maintained, we are able to perceive, and distinguish between, straight lines, curved lines, squares, circles, and the like. It is true that under ordinary conditions the eyes are always moving to a small extent, but these tremors, drifts, and small saccadic corrections do not seem to be at all associated with the contour that is perceived. If efference activated by visual input is important in visual perception, something other than actual efferent output from the central nervous system to the extraocular muscles must be involved.

The few who have suggested, or attempted to formulate, a theory of visual perception in terms of efference have, of course, recognized this and have maintained that "readiness" to issue efferent instructions is the basis for visual perception. Thus, Breese (1899), trying to explain the fluctuations obtained when there is binocular rivalry, states: "consciousness arises only when the cortical centers involved are *ready* to discharge toward the periphery [p. 60, italics ours]." He does not, however, attempt to specify what he meant by such a state of readiness.

More recently, Sperry (1952) has also suggested such a view:

> If there be any objectively demonstrable fact about perception that indicates the nature of the neural process involved, it is the following: Insofar as an organism perceives a given object, it is prepared to respond with reference to it. This preparation-to-respond is absent in an organism that has failed to perceive. . . . The presence or absence of adaptive reaction potentialities of this sort, ready to discharge into motor patterns, makes the difference between perceiving and not perceiving [p. 301].

Sperry, however, is no more specific than Breese (1899) about this suggestion. He simply made statements such as ". . . the preparation for response *is* the perception [p. 301]" and ". . . perception is basically an implicit preparation to respond [p. 302]" and urges the possible value of such a theoretical approach.

Very recently, Taylor (1962) proposed a more elaborated theory based on this same kind of idea. On the basis of his evaluation of existing data he came to the conclusion that all visual perception is learned. In facing the question of exactly what it is that the person learns, and how he learns it, Taylor stated that as a result of appropriately reinforced experience the person learns the appropriate motor responses to make to precisely given constellations of stimulus input.

So far, of course, this is not a very radical suggestion. He proceeded, however, to propose that the conscious experience of visual perception is nothing more or less than these learned responses. Specifically, Taylor developed a system in which, over a large number of repeated trials with appropriate reinforcement, the person learns to make a given response, say an eye movement or a hand movement, to a given visual input, taking account almost automatically of eye, head, and body position. The result of this learning is the formation of "engrams" that may be regarded as well-learned response tendencies that are triggered off by the visual input. These engrams, when they become well established, are automatically brought into play by the appropriate stimulus input. The totality of the engrams that are activated at any moment *is*, for Taylor, the conscious experience of visual perception. In short, Taylor said that what the person "sees" are the readinesses to respond that, over many years, he has learned.

Perhaps, to be complete, one should mention a few others, who in their theoretical considerations give some role to efferent output in affecting perception but not quite in the same way. Hebb (1949), for example, attributed considerable importance to eye movements in learning to perceive contour and shape and, hence, in establishing the cell assemblies that provide the perception of, say, a triangle. Thus, by implication, efference is important in perception for Hebb but he does not spell out any of these implications. Von Holst (1954) also stressed the importance of efference for one particular type of perception. He proposed that the organism is able to distinguish between self-produced movement and externally produced movement by matching a record of the efferent instructions issued to the musculature with the resulting afferent input. Von Holst, however, showed no intention of giving efference a role in perception generally.

Von Holst's theory of "reafference" has been pursued further by Held (1961). Held proposed that, as a result of experience, there is stored somewhere in the central nervous system a set of correlations between efferent output and reafferent input. Because of the wide variety of invariant relationships that provide redundancy through experience, this collection of correlations becomes well established. In a mature organism, following any issued efference, there is an expected reafferent input that should match with what is stored in this correlator. If in any situation the reafference does not fit the expected afferent input, that is, if there is unusual reafference, then there will be some kind of perceptual or behavioral change that occurs. Held was not very explicit about how this occurs except that such experience starts changing what is stored in the correlator. What Held would say about the conscious experience of visual perception is quite unclear.

Some Data about Visual Perception

The fact that a few people over a long period of time have suggested that efferent readiness activated by afferent input is responsible for the conscious experience of visual perception may have no more status than as a curiosity. Perhaps it is more important to note that only a very few have proposed such a view. In order to decide whether or not these suggestions should be taken seriously, we should look at the data that led these persons in this direction. If the data are compelling, that is, if they are difficult to explain in other ways, then these theories should seriously be examined.

Breese (1899) was led to this view by data he collected on binocular rivalry. He presented to one eye of his *S*s a red square with five diagonal lines running from upper left to lower right. On the corresponding part of the retina of the other eye was presented a green square with five diagonal lines running from lower left to upper right. Under such circumstances, as is well known, there is fluctuation of what *S* sees—the red square and the green square alternate in conscious experience. Breese investigated some conditions that affected the length of time that one or another of the two squares was seen. He reported that an effort to pay attention to, say, the red square and keep it in consciousness was effective in increasing the amount of time it was seen only if eye movements occurred when the red field was seen and the eye was relatively still when the green field was seen. If *S* was trained not to make eye movements, effort of will had no effect on the fluctuations. Breese also found that if *S* was instructed to move his eyes along the lines of one of the squares or to count the lines of one of the squares, that square remained much longer in consciousness.

It is easy to understand why such data led Breese to adopt a theory involving efference in the conscious experience of perception. It is difficult to explain the efficacy of eye movements in other ways. If the eyes move when the red square is "seen," the same movement is occurring for both the red and green squares on their respective retinas; if the eyes are relatively still when the green square is "seen," the same relative stillness applies to each square on its respective retina. This suggested to Breese that an explanation should be sought in terms of readiness for motor activity.

Rather different considerations led Taylor (1962) to think in terms of an "efferent readiness" theory of visual perception. The reports of Kohler (1964), originally published in 1951, concerning dramatic changes in the visual perception of contour were of primary importance for Taylor. Kohler had *S*s wear spectacles containing wedge prisms for prolonged periods of time. Such wedge prisms (with bases mounted laterally), among other things, make straight vertical lines appear curved. The dramatic nature of the visual changes that can occur is illustrated by the following quote from Kohler (1964) concerning one of his *S*s:

> After ten days of continuously wearing the spectacles, all objects had straightened out and were no longer distorted. The subject then removed the spectacles. Immediately, impressions of curvature, distortions, and apparent movement set in. The subject complained: "What I experienced after I took off the spectacles was much worse than what I experienced when I first started wearing them. I felt as if I were drunk." Aftereffects continued for four days [p. 34].

In other words, after 10 days of wearing the spectacles *S*'s visual perception had completely adapted and the distortions were eliminated. Curved retinal images were then seen as straight. On taking the spectacles off, distortion, of course, appeared since now straight retinal images were seen as curved. While Kohler did not report such complete adaptation for all of his *S*s, the fact that some *S*s largely, or completely, altered the visual perception of contour is important and requires understanding. How does it happen that the same pattern of retinal stimulation that at one time is "seen" as straight comes to be "seen" as curved?

The observations that Kohler made are not isolated observations. The same phenomenon was reported by Wundt (1898), by Gibson (1933), and more recently by Pick and Hay (1965). Gibson attempted to explain the change in perception of curvature by positing a tendency toward visual "normalization." He discovered that there was some small change toward perceiving less curvature in a contour after simply staring at that contour for 5–10 min. He erroneously assumed that this "Gibson effect" accounted for the entire phenomenon. It has since been shown by Held and Rekosh (1963) and by Cohen (1963) that there is adaptation to prismatically induced curvature over and above the small magnitude involved in the "Gibson effect" and that relevant motor movements are necessary to produce this visual adaptation.

Taylor (1962) felt that these facts forced him to the theory he proposes in which, in our own terms, the efferent readiness activated by the afferent visual input determines the conscious experience of perception. Thus, for him, the perception of contour changes because, while wearing the prism spectacles and engaging in normal activities, *S* learns to make, and then is ready to make, different motor movements in response to the visual input. It is also possible, of course, to think of the changed perception of contour as due to a recoding of the visual input in the central nervous system. But even if one thinks of it this way, one must say that this recoding was determined somehow by the motor activity of the person while he was wearing the spectacles. Perhaps Taylor's position is simpler and more adequate.

Taylor also attempted to derive additional testable propositions from his theoretical statement and to marshall data in support of it. The relationship between his theory and his data will leave many dissatisfied, however. For example, he reasons that, if his theory is correct, the specific visual changes occurring for a person should be unique, depending upon his previous learning and the specific motor adjustments he must make while he is wearing distorting lenses. He documents this by describing his own experiences with adaptation to prismatic distortion —a set of experiences which seem consistent with what he says but are not very compelling.

The most interesting, and most theoretically relevant, data that Taylor presented concern his experiences adapting to prismatically induced curvature when the prism was mounted on a scleral contact lens rather than in a spectacle frame. He realized that if the prism was mounted on a contact lens, the eye itself must move according to the objective contour when scanning that contour and not according to the retinal image. He reasoned that, under these circumstances, adaptation to the curvature distortion should occur as a result of eye movements alone. The *S* would learn quickly to make a different set of motor movements with his eye in response to a given visual input. Specifically, the input to the retina which used to activate "engrams" to move the eye over a curved path would, after some experience with the contact lens, activate engrams to move the eye over a straight path. Once this happened, the perception should have changed since it is based on the evoked "engram" and *S* would see the contour as straight in spite of the curved retinal image. Taylor reported that, indeed, with a prism mounted on a contact lens he adapted to the curvature quickly and completely by just looking back and forth along a contour.

Some data do exist, then, that encourage, even if they do not compel, a theory of visual perception in which the efferent readiness activated by the visual input determines the conscious experience of perception. These data that we have discussed were concerned with the question of "what a person sees." If, however, the conscious experience of visual perception is determined by efference and efferent readiness, we might expect to find some relevant data addressed to the question of "whether or not a person sees." The kind of theory we are discussing here would imply that if, somehow, a situation were created in which the person completely stopped being ready to react to visual input, there might be a cessation of the conscious experience of visual perception. There are two situations that have been used in experiments on visual perception that might produce such a state of affairs, namely, stabilized retinal images and ganzfelds. A close examination of the data reported from such experiments may be useful.

With suitable optical arrangements, a pattern of retinal stimulation can be maintained at a given location on the retina in spite of any eye movements that occur. These images have been called stopped or stabilized retinal images. The experimental findings are that when a retinal image is stopped, contours and shapes tend to disappear. The interpretation of these findings has been in terms of fatigue or satiation of neural mechanisms due to the constant stimulation of the same nerve endings on the retina. The conclusion has been that the ordinary small nystagmic eye movements are essential in maintaining visual perception.

One may also look at the situation produced by a stopped retinal image in another way. There has been a complete destruction of the usual correlation between eye movements and movement of the image across the retina—no matter what eye movements occur, the retinal position remains unchanged. We might well imagine that in such a situation, where movement of the eyes is completely irrelevant to position of retinal stimulation, the person might soon cease responding, or even being ready to respond, to the visual input. Why should the person continue to be ready to make motor movements when these motor movements are useless and irrelevant? From this point of view we might also expect the disappearance of contour and shape with a stopped retinal image as soon as the person stops being ready to react to the input. Furthermore, from this point of view one might expect to observe the same kinds of disappearance of contour even if the stabilized image were not *stopped* on the retina. That is, the stabilized image could be moved, by *E,* across the retina while maintaining the complete lack of correlation between position on the retina and eye movements. Since, from this viewpoint, the important factor would be the cessation of reactivity due to the total lack of correlation between eye movement and movement across the retina, we would still expect the disappearance.

Along these lines an interesting observation is reported by Campbell and Robson (1961). They studied stabilized images produced by making visible the shadows of the retinal capillaries, a very precise way of producing a stabilized image. They reported as follows:

> New findings are that a stabilized shadow of the retinal capillaries disappears in a few seconds and does not reappear even in flickering light. The capillary shadow can be seen for a much longer period if moved across the retina at certain amplitudes and frequencies but, even so, these moving shadows also ultimately disappear and never reappear again spontaneously. Similar observations have been made using the central details of the shadow of the macular pigment [p. 12P].

Clearly, these observations cannot be explained entirely in terms of fatigue or satiation of neural mechanisms. They do fit what we would expect from our conjectures about the importance of efferent readiness in visual perception.

There is another known situation in which persons experience the cessation of visual perception. This occurs sometimes if a person's total visual field is a "ganzfeld," that is, a completely homogeneous, structureless field of vision. Typically, when viewing a "ganzfeld," *O* perceives a fog or mist of light and does not perceive any surface. Cohen (1960) reported that about one-third of *S*s in his study also experience "blank out," that is, "complete disappearance of the sense of vision for short periods of time." It is conceivable that, in the absence of any structure in the visual field that the person can use for fixation, the person occasionally stops responding altogether to the visual input. We will not pursue the speculation further. We will leave it at this except for one piece of data that is interesting with respect to this speculation. Tepas (1962) found that there was a significant absence of saccadic eye movements just prior to the onset of a "blank out." The absence of saccadic eye movements continued during the blank out and the end of the blank-out period coincided with the resumption of such eye movements.

Some Theoretical Specifications

If one is to take seriously a theory proposing that the conscious experience of visual perception is determined by efferent readiness activated by afferent visual input, it is necessary to specify some of the characteristics that such a theory must have to fit the known data. It is also necessary to specify something about what "efferent readiness" is, how it is developed, how it is activated, and what particular efferent readinesses affect visual perception.

The published literature discussed above suggests that the visual perception of contour can be altered. If we wish to say that this visual perception is determined by efferent readiness, then we must postulate that the efferent readiness appropriate to a given visual input must, to at least some extent, be learned and modifiable. This would make it likely that visual perception of contour and shape must be learned, or at least in part. We might imagine that the visual experience of a newborn infant has no sharp contours or definite shapes but consists entirely of fuzzy blotches of brightness and color differentials. Perhaps more than this is innately built into the organism but we need not concern ourselves with the problem. Even if only this much is built in, mature visual perception based on efferent readiness can easily develop.

There is a precise and invariant relationship that always holds between the magnitude and direction of an eye movement and the magnitude and direction of movement across the retina of any point of stimulation on the retina. Thus, of course, the basis exists for being able to learn the appropriate efferent instructions to issue to move a point of stimulation from one part of the retina to any other part. To learn this, however, would mean to learn an almost countless number of efferent sets of instructions and the human organism almost certainly does not learn all of this. The part of the retina that is of the greatest interest is the fovea. We can imagine a tendency in our newborn infant to examine, in detail, parts of the visual field in which the brightness or the color changes. When such parts of the visual field fall on the fovea, the detail is best and so the organism learns to direct the eye so as to bring points of peripheral stimulation to the fovea. This, of course, is a much more restricted and manageable set of efferent instructions to learn. Most adult organisms have never learned to, and cannot, execute eye movements to bring a point of stimulation from 20 degrees to the right of the fovea, for example, to a point 10 degrees below the fovea. Thus, for example, Fender (1964) reported that "subjects find it almost impossible to track a moving target while maintaining fixation a few degrees away from it [p. 315]."

It is, then, a manageable set of efferent instructions that the organism learns to issue for eye movements. Considering that the saccadic fixating eye movement in the adult is not a completely accurate movement—errors up to half a degree are not unusual—the amount to be learned is manageable indeed. Thus, we can imagine the input to the retina coded as if the retina were calibrated in terms of distance and direction from the fovea. The organism learns the appropriate efferent instructions to be issued from the central nervous system to direct the eye to move so as to bring any point of stimulation on to the fovea. After a considerable amount of learning has gone on, these sets of efferent instructions can be viewed as becoming "preprogrammed" and as being automatically activated by brightness or color differentials stimulating the retina. We do not mean the term "activated" in the sense of efference actually being sent from the central nervous system through the motor neurones; we mean to use the term in the sense that these preprogrammed sets of efferent instructions are brought into a state of immediate readiness for use. Thus, those efferent instructions which if issued would bring areas of brightness differential onto the fovea, are ready for immediate use.

Although the idea of efferent instructions held in readiness for use is not a new one, as we have seen, it is still a rather vague one. Some attempt at clarification would perhaps be helpful. Without trying to speculate about the exact physiological mechanisms and arrangements, it seems plausible to imagine that the physical system is limited in the number of sets of stored preprogrammed instructions that can be "immediately" sent out through the motor pathways. Thus, out of the very large number of sets of efferent instructions that the organism has learned, only some are held in readiness for immediate use. Without intending any precise analogy, we could imagine a jukebox which, at the push of the appropriate button, will immediately play any of a hundred different phonograph records. The owner of the jukebox could also have many thousands of other records available but, obviously, they cannot be played immediately. The owner could, however, with a little bit of work, change the entire set, or part of the set, of the hundred records that are immediately available for playing. One might think of these hundred records as being "ready for immediate use."

If we are to think of the conscious experience of visual perception as being determined by these preprogrammed sets of efferent instructions that are activated into readiness by the afferent input, then it is necessary to specify something about the level of generality or specificity of the efferent signals that are issued from the central nervous system. If, using an efference readiness theory of visual perception, we are to be able to have a person perceive a given shape or contour as the same no matter what the eye position is when viewing it, it seems necessary to specify that the efferent instructions issued from the central nervous system must be general in nature, that is, relatively far removed from the final signal that causes the exact muscle twitch. Thus, the efferent signal from the central nervous system could be concerned only with direction and magnitude of deviation from the fovea, and final computation to effectuate the actual muscle contractions could take account of afferent information at more peripheral levels. This is not an unlikely state of affairs. For example, Merton (1964) said: "Hughlings Jackson (no reference given) showed that, in the code used by the motor cortex, the orders sent out represent instructions to perform movements, not instructions to individual muscles to contract [p. 399]."

One might be tempted to say that the conscious experience of visual perception of contour and shape (and similar arguments could be made for perception of distance and depth) was determined by readiness to issue efferent instructions to the extraocular muscles. However, it is clear that this cannot be the whole story. The data show clearly that fairly large changes in the visual perception of curvature occur if wedge prism spectacles are worn for long periods of time. Pick and Hay (1964), for example, found an average of 30% adaptation to curvature in eight *S*s who wore such spectacles all their waking hours for 42 days. But if *S* wears prism spectacles, no change in the efference or efferent readiness activated by the visual input should occur with respect to the extraocular muscles. The eye, under these conditions, must move with respect to the retinal contour, not with respect to the objective contour. However, head movements, arm movements, and all other body movements must, in order to be effective, correspond to the objective contour and so, efference and efferent readiness concerning these motor movements that are activated by the visual input must change. If we are to explain these adaptations to curvature in terms of an efferent readiness theory, then, it is necessary to say that the conscious experience of visual perception is determined by the total efferent readiness activated by the visual input, not just the efference relevant to the eyes.

Perhaps this represents sufficient specification of an "efference readiness theory" to permit submitting the theory to experimental test. Before proceeding to examine this question, however, it is necessary to consider an alternative interpretation of the data on which we have relied in the discussion. Harris (1965) proposed that the visual system and visual perception are probably not changeable and that the changes that occur when adapting to distorting spectacles are proprioceptive changes, that is, the end result of the adaptation is a change in the felt position of some part of the body. He stated,

> Vision seems to be largely inflexible, whereas the position sense is remarkably labile . . . proprioceptive perception of parts of the body (and therefore of the location of touched objects) develops with the help of *innate visual perception* . . . [italics ours, pp. 441–442].

Harris presented data to show that such changes in "felt position" do indeed occur as a result of adaptation to displacement of the visual field. He also presented a highly ingenious analysis in terms of changes in "felt position" to explain adaptation to spectacles that invert or reverse the visual field. With regard to curvature, Harris implied (although he retracted the implication partly in a footnote) that the same end result is achieved rather than any change in visual perception. He said,

> So perhaps adaptation to curvature also involves altered registration of eye movements without any change in scanning behavior. After adapting, the subject may feel that his eyes are moving in a straight line when they are actually tracing out a curve [p. 428].

On examination, however, this suggestion seems strange and even rather inconsistent with the position taken by Harris. First of all, if *S* is wearing prism spectacles, what could conceivably lead to a change in felt position of the eyes? The eye movements necessary to scan a contour would be precisely consistent with vision (retinal input) which is "largely inflexible" and "innate." After moving about in the environment an *S* might be expected to recalibrate the felt position of other parts of the body but *not* of the eyes. Second, Harris seemed to be making a very curious suggestion. He was apparently suggesting that the visual perception of a contour such as curvature is determined by how *S* feels his eyes are moving when he scans this contour. It becomes unclear what Harris meant by visual perception. Did he suggest that with steady fixation contours cannot be perceived, or did he suggest some notion of efferent readiness similar to ours?

Possible Experimental Tests

If the conscious experience of visual perception of contour is, indeed, determined by the efferent readiness activated by the visual input, there is a definite empirical implication that can be tested experimentally. If, without changing anything about the pattern of retinal stimulation, one could alter the particular preprogrammed sets of efferent instructions that were activated and held in readiness for immediate use, one would expect to produce a change in visual perception.

It is clear that, by using prism spectacles that produce a curved retinal input when looking at a straight line, one could induce a person to learn, say, to make a straight arm movement in response to a curved retinal image. It is also possible, as Taylor (1962) pointed out, that by using prisms mounted on contact lenses one could induce a person to make a straight eye movement in response to a curved retinal image. The problem is how to do these experiments with appropriate controls and appropriate comparison conditions so that alternative interpretations of the data can be ruled out. Since a major class of possible alternative interpretations of such changes in visual perception would be based on the idea that the change occurs because of conflicting information obtained from retinal input and from feedback from the muscles and joints, it would seem to be desirable to control for these factors. In other words, we would want different experimental conditions in which retinal input was identical and feedback from muscles and joints was identical but the efference issued from the central nervous system was different.

Theoretically, of course, it is possible to have quite different general efference issued ending in exactly the same muscle contractions. To use an analogy, instructions could be issued to a computer to divide 28 by 4 or to find the fifth prime number. As a result of these two very different instructions, the machine ends up doing the identical thing, namely, it prints the number 7. Operationally, it does not seem easy to do but perhaps one may approach such a situation. For example, if an *S* wearing prism spectacles looks at a straight edge that appears curved and is instructed to run his finger along the edge, pressing on the edge, the finger will actually move in a straight line, of course, and feedback from the skeletal joint receptors will provide this information. There is no necessity, however, for efference from the central nervous system to concern itself with the exact contour involved. If, however, *S* had to learn to make an accurate sweeping motion with his finger that corresponded to the contour without any edge to press on, it seems more likely that the efference from the central nervous system would have to concern itself with contour. In both conditions, however, the feedback from the limb would be very similar. Only in the latter conditions would we expect *S* to develop the new efferent readiness to move his arm in a straight path when the retinal input is curved and, hence, only in that condition would we expect change in visual perception.

Four experiments have been designed to provide such tests of the theory that the conscious experience of visual perception is determined by efferent readiness activated by the visual input. They all used, for this purpose, the empirical vehicle of adaptation to prismatically induced curvature.

EXPERIMENT I

In this preliminary experiment an attempt was made to create two experimental conditions similar with respect to the active movements that occur and the proprioceptive feedback but different with respect to whether or not *S* learns a new afferent-efferent association. One-half of the *S*s were required to learn to move a stylus in a continuous movement along a path between two brass rods. Since the path between the rods either was objectively straight and appeared curved or else was objectively curved and appeared straight, *S*s needed to learn a new, unitary, afferent-efferent association in order to perform the continuous tracking motion without error, i.e., without striking one of the rods. The remaining *S*s were not asked to learn a continuous tracking motion. Rather, they were asked to move the stylus, as slowly as they wished, between the rods with the paramount objective of never touching a rod. Thus, one-half the *S*s were asked to learn a new, unitary, skilled movement; the remaining *S*s were not.

Method

Subjects.—The *S*s in this experiment were freshmen or sophomore female students at either Stanford University or Foothill College. All were right handed and did not wear spectacles. Each was paid $4.00 for participating in the experiment. Only females were used in the study because preliminary work had indicated that males tended to become more frustrated at the boring nature of the task and tended to lose interest and motivation. The *S*s were run, assigned to experimental conditions at random, until 10 usable *S*s in each condition were obtained. During the course of conducting the experiment, the data from 12 *S*s were discarded for the following reasons: (*a*) Three *S*s because of difficulties with the biteboard during the session; (*b*) Two *S*s because of disregarding the instructions; (*c*) Seven *S*s because of highly inaccurate or suspicious initial settings with the prisms.

Apparatus.—The main piece of apparatus was a white formica board, 40 in. wide × 26 in. high. The board was held vertically in a wooden frame and rested on a table. Down the middle of the board ran two parallel vertical brass rods. The rods were slightly less than $\frac{1}{8}$ in. in diameter and were mounted $\frac{1}{2}$ in. apart between centers. The ends of the rods ran through holes at the top and bottom of the board. The ends of the rods were free to slip back and forth through these holes. The midpoints of the rods were rigidly attached to a horizontal center strip in the board which had an identical surface set flush with the surface of the rest of the board. A knob was mounted in the frame below the board and to the right of center. By turning this knob, *S* could move the center strip back and forth and thus adjust the rods to various desired degrees of curvature. On the back of the board was a pointer which ran along a centimeter scale. The pointer measured the horizontal deviation of the midpoints of the rods from true straight. A biteboard was mounted on the table directly in front of the center of the board so that the distance from *S*'s eyes to the board was approximately 40 cm.

During most of the experimental session *S* wore goggles with 25-diopter prisms mounted with their bases left. While wearing the goggles and biting on the biteboard, the vertical extent of vision on the board was about 37 cm. For measurements involving the naked eye, *S*s wore a similar pair of goggles with plate glass in them. For each *S*, the height of the chair and of the biteboard was adjusted so that her eyes were at the same height as the horizontal center strip on the board.

Procedure.—Before any goggles were put on *S*, she was shown how, by turning the knob, the curvature of the lines could be changed. She was told that periodically she would be asked to adjust the lines so that they were straight. She was also told that she would spend part of the session moving a stylus down between the two rods using her right hand and was shown that if the stylus touched either rod a buzzer would sound. The type of stylus stroke was then explained to *S*, the specific instruction depending on the experimental condition.

In the condition designed to encourage learning a new afferent-efferent association (Learning condition) *S* was told to make a smooth, fast, sweeping motion with the stylus between the two rods. She was told to try to learn to avoid hitting the rods but not to be concerned about hitting them at the beginning. She was not to slow her motion down in order to avoid hitting the rods but to continue a smooth, rapid motion and gradually improve her performance. The smoothness and rapidity of the stylus stroke and the objective of learning to make the stroke better were emphasized.

In the condition designed to minimize the learning of a new afferent-efferent association (Accuracy condition) *S* was told to make a slow, very careful movement of the stylus between the rods so that she would not hit either rod. She was told to move slowly enough to be sure she was accurate. The importance of going slowly and never hitting the rods was stressed.

It was intended that, in the Learning condition, *S*s would have to learn to make a straight arm movement when the retinal input was curved or a curved arm movement when the retinal input was straight. To the extent that they learned this, a new efferent readiness would be activated by the visual input. It was also intended that, in the Accuracy condition, *S*s would respond primarily to the local deviation of the stylus from the rod and would never learn anything new about efferent instructions to the arm relevant to the contour. In order to keep the amount of experience constant for *S*s in the different experimental conditions, each was instructed that a bell would sound every 12 sec. At this signal she was to insert the stylus between the two rods at the top of her visual field and make the downward stroke, ending near the bottom of her visual field. Thus, each *S* had exactly the same number of stylus strokes and the same time spent looking at the lines.

After telling *S* that, from this point on, she was to have her eyes open only while biting on the biteboard, she was asked to shut her eyes and *E* put the plain glass goggles on her.

The *E* then moved the rods a few centimeters to the left of where they would look approximately straight and *S* was asked to turn the knob so as to make the lines straight. The setting was recorded to the nearest ½ mm. He then moved the rods off in the opposite direction and another setting was made. This was continued until four measurements were obtained. The *S* then closed her eyes and the plain glass goggles were replaced by prism goggles. Initial settings of straight with the prism goggles were obtained in a similar manner.

The *S* closed her eyes again while *E* set the lines at the proper position. If *S* was in an "apparently straight" condition, the rods were positioned at the average of the settings of "straight" that *S* had just made wearing the prism goggles. If *S* was in an "apparently curved" condition, the rods were positioned at the average of the settings of straight that *S* had made wearing the plain glass goggles.

The *S* was then asked, with her eyes shut, to run her fingers up and down the two rods until she could tell whether they were straight or curved and, if they were curved, in which direction. The purpose of this was to provide some information to *S* that might help in the performance of the task. This aspect of the procedure was probably unnecessary. It was omitted in Exp. II reported below.

The *S* then opened her eyes, took the stylus in her hand, and after *E* quickly reviewed the stroking instructions and started the bell, began the actual practice. There were five stroking periods each intended to be 10 min. long. Some of the periods for *S*s in the Accuracy conditions were longer since, if they skipped some of the bell rings, the period was extended so that there would be 50 strokes in each period. The stroking periods were separated by rest periods of 3 min. during which *S* leaned back with eyes closed.

During the stroking periods, *E* observed the speed of *S*'s strokes. If *S* in a Learning condition stroked too slowly (more than 1 sec. per stroke), she was reminded to go faster. The *S*s in the Accuracy conditions were reminded to slow down if they went too fast (less than 4 sec. per stroke). The *S*s in the Accuracy conditions were also reminded to be careful if they hit the rods, telling them to go slowly enough so that it would not happen.

Following the fifth 10-min. stroking period, while *S* rested, the board was washed to remove the slight traces left by the stylus. The *S* then spent 2 min. stroking and immediately afterwards the final settings of straight while wearing the prism goggles were made. The rods were then returned to the stroking position for that *S* and she stroked for another 2 min. With *S*'s eyes shut, *E* removed the prism goggles and put the plain glass goggles in their place. He quickly washed the board off again and had *S* open her eyes and make the final settings looking through plain glass.

Results

Initial straight settings.—Table 1 presents, for each of the four experimental conditions the average initial setting of straight with the naked eye (plain glass goggles) and with the prism goggles. It also presents the average change from the beginning to the end of the experimental session for each of these measures.

There are only minor variations among the four experimental conditions on the initial measurements. The average setting of straight with the naked eye is very close to objectively straight (between 9.90 and 9.95 on our measurement scale). The average setting of straight with the prisms varies slightly around 4.40; in other words, the prisms produced a curvature of about 5.5 cm. displacement of the middle of the line. Analysis of variance on the initial measurements revealed that none of the differences among experimental conditions even approached statistical significance.

Change from initial to final measurements.—These showed systematic differences in line with what one would expect on the basis of an efferent readiness theory. Although the effects were small, they were reasonably consistent. On both the measures of adaptation (changes measured with prisms in place) and aftereffects (changes measured with the naked eye) the Learning condition yielded more changes in visual perception than the comparable Accuracy condition. Neither of these quite reached conventional levels of acceptable significance.

We can increase the reliability of our measure of change of visual perception, however, by simply averaging for each *S* the adaptation and aftereffect measured. An analysis of variance on this combined index yielded $F(1, 36) = 6.46$, $p < .05$. Thus, we may conclude that the Learning conditions produced more change than the Accuracy conditions.

The differences between the apparently straight and apparently curved conditions were, of course, highly significant in all cases. This difference was due to the operation of the "Gibson effect" in the Apparently curved conditions and its absence in the Apparently straight conditions. It is clear, however, that the difference between the Learning and the Accuracy conditions existed independently of the Gibson effect.

Discussion

The data were consistent with the implications from an efferent readiness theory of visual perception of contour. In the condition intended to force *S* to

TABLE 1

INITIAL MEASUREMENTS AND CHANGES (IN CENTIMETERS) IN THE PERCEPTION OF A STRAIGHT LINE (EXP. I)

	Experimental Cond.			
	Apparently Straight		Apparently Curved	
	Learning	Accuracy	Learning	Accuracy
Initial with Prisms	4.55	4.34	4.29	4.39
Change with Prisms	+.28	+.10	+1.59	+1.31
Initial with Naked Eye	9.92	9.96	9.90	9.96
Change with Naked Eye	+.18	+.02	+.86	+.65

learn a new afferent-efferent association significantly more change in the visual perception of curvature was obtained than in the condition intended to make such learning unlikely. Whether or not the Learning and Accuracy conditions really had their intended effect is, of course, not directly answerable from the data. All one can say is that the results obtained are in line with the predictions made from the theory and from intuitive notions as to the effects of the experimental manipulations.

The magnitude of the effect created by the experimental manipulation is, clearly, very small, being about 2 mm. between the Learning and Accuracy condition. It is, of course, unclear as to whether this magnitude of change of visual perception is, or is not, disappointing. Considering that visual perception is very likely heavily dependent upon efferent readiness concerning the extraocular muscles, and there was certainly no change in these afferent-efferent associations, and considering that the only other movement involved at all was that of the arm, one might not expect a very large change in visual perception.

The data have relevance to Held's (1961) theory concerning the importance of reafference. Both the Learning and the Accuracy conditions are, in Held's sense, active movement conditions. In both conditions the afference, and hence the reafference, would be unusual and from Held's theory one would expect perceptual change equally for both. It seems clear from the data, however, that Held's distinction between active and passive movement was too gross. Distinctions have to be made concerning the specific nature of the efference.

EXPERIMENT II

The attempt was made in the design of the previous experiment to keep the proprioceptive input from the arm the same in both Learning and Accuracy conditions so as to rule out possible interpretations of the visual changes as having been due to recoding of visual input based on information obtained from proprioceptive input. For this reason there were two rods spaced closely together so that, since the stylus was confined between the two rods in all conditions, the informational input from muscle and joint receptors would, of necessity, be similar for the two conditions. However, *E*s were obviously not successful in making the proprioceptive input identical in these experimental conditions since, at a minimum, the rate of input was systematically different. In the Learning conditions the hand moved quickly while in the Accuracy conditions the hand moved slowly. Perhaps the rate of informational input is important. In this experiment, consequently, a theoretical replication with quite different instructions to *S* was attempted so that the rate of input, as well as the specific proprioceptive information, would be held constant.

Method

Subjects.—Sixty-two females, 22 freshmen or sophomores from Stanford University and 40 junior and senior high-school students, participated in the experiment. All *S*s were naive about the experiment and were paid $3.00 for participating. The *S*s were randomly assigned to one of four experimental conditions with the restriction that the ratio of college to high-school *S*s be kept nearly equal. The data from two *S*s were discarded—one because she did not follow the instructions and the other because the apparatus failed during the experiment.

Apparatus.—The apparatus for the experiment was identical to that used in Exp. I except for a few minor changes. A sturdier biteboard was constructed; movable clips were attached to the rods at the top and bottom of *S*'s visual field so as to be sure that the stylus movement was always entirely within the visual field. The *S*s were seated so that the distance between their eyes and the rods was about 48 cm., approximately 8 cm. farther away than in Exp. I.

The experiment was conducted monocularly throughout so that we would also be able to measure interocular transfer. For this reason three sets of goggles instead of two were used. One set contained a 25-diopter prism mounted base left in front of the right eye, the left eye being occluded; the other two sets had plain glass, one in front of the right eye, the other in front of the left eye. The other eye was always occluded. There were two viewing conditions, one in which *S* viewed an objectively straight, apparently curved line; the other in which she viewed an objectively curved, apparently straight line. There were two movement conditions, one in which we attempted to maximize the learning of a new afferent-efferent association and one in which we attempted to minimize such learning while holding other variables constant. The instructions to *S*s in the Learning condition emphasized the learning of a smooth, fast, sweeping motion of the stylus between the rods. The *E* demonstrated the stroke and the buzzer sound resulting when a rod was touched by the stylus. The *S* was told that touching the rods was to be accepted at first, that the important aspect of the task was to learn the smooth stroking motion required to move the stylus between the rods. In the Contact condition *S*s were instructed to learn a smooth, stroking motion while maintaining pressure on one rod. The *S*s in both conditions were encouraged to rest when this was needed and to proceed at the task at a self-determined pace.

Instructions were also given concerning how to set the line so that it was straight and to keep the eyes closed any time *S* was not biting on the biteboard. Measurements were made in the same way as in Exp. I with the addition that separate measurements were taken for each naked eye at the beginning and end of the experiment.

During the experimental period *E* recorded the number of strokes made and the cumulative time on and off the biteboard. He also recorded any verbal reports given by *S* while she was leaning back and resting. If *S* was not following instructions, she was corrected at once. When the time on the biteboard had accumulated to 10 min., 20 min., and 30 min., *E* reminded *S* of what she was supposed to be doing; e.g., "You're doing fine, but let me remind you that the important thing is that you make firm contact with one of the rods," or "You're doing fine, but let me remind you that you're supposed to be trying to learn to make a fast, smooth, sweeping stroke."

When *S* had been on the biteboard for 40 min., she was asked to close her eyes and to lean back. At this point *E* removed from the board any traces left by the stylus. The *S* was then asked to work for a little while longer. After *S* had stayed on the biteboard for 45 sec., she was instructed to remain on the biteboard but to close her eyes. Any traces of the stylus were again removed and the final measurements of straight were taken.

At this point *E* informed *S* that the experiment was over, but that he would like to ask a few questions. The *S* was asked how the board looked, whether she noticed any changes in the curvature of the rods, and how her eyes felt during the experiment.

Results

Table 2 presents the data for each experimental condition. As in Exp. I, the differences among the four conditions in the initial measurements were very small. In all four experimental conditions, the average settings of straight were slightly under 10.0 on the measurement scale with the right naked eye. With the left naked eye they were a shade over 10.0. This difference was significant; 52 *S*s showed a higher mean setting with the left naked eye than with the right naked eye; 7 *S*s showed a difference in the other direction; and for 1 *S* the average settings were equal. This difference was undoubtedly due to the slightly different angle of view between the two eyes. The average setting of straight with the prism spectacles was 4.96, a curvature represented by about 5 cm. displacement of the rods from the middle of the board.

The data on changes from initial to final measurements were very similar to those obtained in Exp. I. In the Learning conditions, where one would expect some learning of new efferent instructions activated by the visual input, greater change in visual perception was obtained than in the Contact conditions. An analysis of variance on the changes of the settings of straight with prisms yielded $F\ (1, 56) = 8.92$, $p < .01$, for the difference between the Learning and Contact conditions. A

TABLE 2

Initial Measurements and Changes (in Centimeters) in the Perception of a Straight Line (Exp. II)

	Experimental Cond.			
	Apparently Straight		Apparently Curved	
	Learning	Contact	Learning	Contact
Initial with Prism (right eye)	5.02	4.86	4.94	5.01
Change	+.23	+.15	+1.20	+.88
Initial with Right Naked Eye	9.77	9.72	9.74	9.83
Change	+.32	+.20	+.91	+.68
Initial with Left Naked Eye	10.04	10.03	10.09	10.10
Change	+.14	+.05	+.35	+.20

similar analysis of variance on the changes from initial to final measurements for the right naked eye (the eye that wore the prism) showed the Learning and Contact conditions to be significant also, $F\ (1, 56) = 5.47, p <$ .05. The combined index of adaptation and aftereffect that was used in Exp. I was, of course, highly significant, $F = 15.57$.

The results for transfer from the right naked eye to the left naked eye were less clear. In all of the experimental conditions there was some transfer to the left naked eye and the amount of change in the left eye was greater in the Learning conditions than in the comparable Contact conditions. The difference between the Learning and Contact conditions was not statistically significant, however, $F = 2.82$. The measurements on the left eye were always taken after the measurements on the right eye in this study. It is impossible to assess the effect of the time delay on the results.

Similar to the results of Exp. I, there was a highly significant difference on all measurements between the conditions in which an apparently straight or apparently curved line was viewed. This Gibson effect clearly was independent of the changes of primary interest.

Discussion

The results of Exp. II completely supported the results, and the interpretation of the results, from Exp. I. In spite of the fact that quite different instructions were used to create conditions that would minimize the learning of new afferent-efferent associations, the results came out in the same way. In Exp. I *E*s depended on an instruction to go very slowly and to avoid ever hitting the rods. It was intended that this would force *S* to concentrate on the local deviation of the stylus from the rod and that she would, therefore, not learn new efference to issue in response to the visual input. In Exp. II *E*s depended, for the same purpose, on an instruction to maintain pressure and contact with one rod during the whole stroke. It was hoped that, since the movement of the arm would thus be guided by the actual rod, *S* need not, and would not, learn a new afferent-efferent association. The results of the two experiments support the interpretation that visual perceptual change occurs if one changes the efferent readiness activated by the visual input.

In both experiments the actual arm movement, and hence the actual proprioceptive feedback from the arm movement, was nearly identical for the Learning conditions and the comparable nonlearning conditions. In Exp. I it was possible to argue that, since the rate of proprioceptive input was different (fast vs. slow movements), perhaps this affected the results. In Exp. II this difference did not exist. The rate of movement was similar in all experimental conditions—if anything, the movement was faster in the Contact conditions. Table 3 shows the number of strokes made on the average by *S*s in each experimental condition in the 40 min. of actual stroking. None of the differences are statistically significant but it is clear that the difference that does exist is in the direction of more strokes per unit time in the Contact conditions. Hence, it is no longer plausible to suppose that the rate of proprioceptive input affects the results.

It is worth pointing out that while the two different kinds of nonlearning conditions did probably reduce the extent to which *S*s learned new afferent-efferent associations, we cannot be sure that these conditions prevented such learning altogether. The data from the Contact conditions in Exp. II provide some basis for assessing whether some learning did occur. One might expect that if learning occurred in these Contact conditions, it would probably depend on amount of experience to a greater extent that it would in the Learning conditions. In the Contact conditions those who made very many strokes might be more likely to have learned some new efference. To examine this possibility we computed the correlations, within each experimental condition, between the number of strokes made and the combined index of adaptation and aftereffect for the right naked eye. These correlations are presented in Table 3. There is no significant correlation at all for the two Learning conditions but significant (at the 5% level) correlation for each of the Contact conditions. It seems, then, that some *S*s in the Contact conditions did learn. If such learning could have been entirely prevented, the difference between conditions would, presumably, have been larger.

EXPERIMENT III

Experiments I and II, while supportive of the theory, have in common a possible confounding factor. In the Learning conditions there were frequent error signals that were absent in the nonlearning conditions. The third experiment was designed to eliminate this possibly confounding factor while testing the theory again under very different empirical conditions.

Two general methodological changes were made in the third study: (*a*) Adaptation resulting from a change in efferent readiness and the adaptation resulting from the "Gibson normalization effect" were experimentally separated by allowing normalization to develop prior to the introduction of arm and hand movements; and (*b*) measurements of adaptation were made after short periods of activity to make it possible to study the course of adaptation throughout the experimental session.

Method

Subjects.—All *S*s were males, either high-school seniors or college students. All had

TABLE 3

NUMBER OF STROKES AND ITS CORRELATION WITH THE COMBINED INDEX OF PERCEPTUAL CHANGE

	Experimental Cond.			
	Apparently Straight		Apparently Curved	
	Learning	Contact	Learning	Contact
Number of Strokes	624.00	689.93	585.33	626.73
r between Adapt + Aftereffect and Number of Strokes	−.172	+.525	+.022	+.500

good, uncorrected vision sufficient for an unrestricted driver's license or had vision fully corrected by contact lenses. A total of 73 *S*s participated in the study but only data from 54 (9 in each condition) were used in the analysis. Sixteen of the discarded *S*s made very inaccurate level settings during initial measures with prism goggles. One *S* was unable to make settings within time limits, and two *S*s used background cues as a basis for their settings after the shooting period.

Procedures.—The technique used in this study to provide *S*s with an opportunity to make arm and hand movements discrepant with their visual input employed a shooting gallery. The *S*s "shot" a pistol emitting a continuous light ray at a target that moved back and forth on a track. While engaging in this activity, they wore prism spectacles. When the light ray hit the center of the target, a photocell and relay mechanism activated a buzzer. In one experimental condition the light was visible; in the other an infrared filter was placed over the barrel of the pistol.

No adaptation to the prism induced curvature was expected in the visible light condition. The efference issued in this condition need not be made in response to the path of the target's movement or the contour of the track but rather to the discrepancy between the seen position of the light ray and the target's position, speed, and direction of movement. In essence, these *S*s could act as servomechanisms as they performed a simple tracking task. Yet their arms and hands would move in a path consistent with the distal contour and be discrepant with the perceived proximal contour.

Ideally, *S*s shooting with the invisible infrared light would have been forced to issue efference activated only by the perceived contour of the target's path. Guided by the information from the buzzer when on target, they would have to learn a new set of efferent responses to the distorted perception. Therefore, they would be expected to adapt to the prism induced curvature. Pretesting with this manipulation, however, rapidly led to the conclusion that it was almost impossible to hit the target; hitting occurred rarely and seemingly by chance. Consequently, *S*s were permitted to aim while shooting with the infrared light. With aiming they were able to hit the target, although still with some difficulty, and make the required arm and hand movements. Although aiming was not necessary for *S*s shooting with the visible light, they were also told to aim in order to equate this factor.

The two conditions involving infrared and visible light were combined with two conditions of viewing an apparently straight or apparently curved line, resulting in a 2 × 2 design. After these four initial groups were run, two supplementary groups of *S*s were added to clarify the findings and interpretations of the results. The *S*s in the Aim-only condition shot with the infrared light and were allowed to aim but received no information as to when they hit the target. The *S*s in the other supplementary condition also shot with the infrared light and received no information. In addition, they were not allowed to view their arms, hands, or the barrel of the pistol. This No-information group was designed to determine whether there were any factors in the experimental situation which would result in a change in contour perception if *S*s neither made discrepant efferent responses nor received any atypical visual reafference. The *S*s in these two supplementary groups viewed only an apparently straight contour.

Apparatus.—As indicated above, the experimental apparatus consisted of a prism to produce a curvature transformation of the visual world and a shooting gallery to give *S*s a means of engaging in activity with the distorted world. In addition, there was a method for measuring adaptation to curvature.

A 30° wedge prism of optical plastic, 4 in. long and $1\frac{3}{4}$ in. in height, was used to produce curvature. It was mounted base upwards in welder's goggles with the front of the prism flush with the outside of the goggles. The field of view through the prism goggles was 86° wide and 48° high. Similar goggles with a plain piece of glass were used when *S*s viewed the same size visual field without any distortion. The shooting-gallery component of the apparatus consisted of a 9-ft. horizontal track across which a target box moved at a rate of 1.5 ft/sec. The reversal and reacceleration of the target box were virtually instantaneous. The actual target was a 1 × 1 cm. photocell which was sensitive to both visible and infrared light. It was mounted in the middle of the target box; a series of concentric red and white rings surrounded the photocell and enhanced its target-like appearance. When the photocell was activated by light, a relay closed, starting a buzzer and clock. The buzzer signaled *S*s that they were on target. The clock provided a record of the amount of time spent on the target.

The *S*s shot at the photocell with a pistol emitting a continuous collimated ray of light approximately 1 in. in diameter. The infrared filter used in the invisible light conditions, inserted in front of the barrel of the pistol, effectively blocked visible light under the illumination conditions used in the study.

The track on which the target ran could be bent into a smooth curve. The track itself was attached to an aluminum bar 9 ft. long, 3 in. wide, and ¼ in. thick. The bar was held at a constant height at both ends by supports and forced up or down by pressure at the middle. A threaded rod affixed with a bracket to the center of the bar was raised or lowered by a motor and pulley arrangement. The position of the bar was measured by a cord running from the threaded rod along the length of a meter stick placed at the front of the table where *E* sat. The position of an indicator on this cord accurately reflected the position of the middle of the bar.

It was impossible to set the bar to appear perfectly straight with or without the prisms. Since it was supported only at the ends and middle, the bar sagged slightly at the ¼ and ¾ points, and curvature produced by bending the bar did not exactly compensate for the curvature induced by the prism. When set to appear approximately straight, the ¼ and ¾ points looked slightly elevated. Consequently, *S*s were always told to set the bar to appear level, with the middle of the bar placed at the same height as the two ends. The *S*s had no difficulty in doing this either with the naked eye or the prism goggles.

The *S*s sat at a table directly in front of the middle of the track with their eyes 64 in. from the bar. From this position the target movement subtended a visual angle of 78°. While making settings, viewing the bar and target, or shooting, *S*'s head was held fixed by a biteboard attached to this table. When *S*s wore the nondistorting goggles, the biteboard was parallel to the surface of the table, and the bar appeared in the center of the field of view. When wearing the prism goggles, the biteboard was angled 15° downward to compensate for the prism displacement effect and make the bar still appear in the approximate center of the field of view.

Black cloth draped irregularly over, behind, and to the sides of the bar blocked *S*'s view of the walls and ceilings of the experimental room and prevented him from realizing that the goggles produced curvature. The threaded rod and motor and pulley arrangement were also hidden by a piece of black cloth to prevent *S*s from using the position of the bar relative to the motor and pulleys as a guide for their settings. In the no-information condition a shield prevented *S*s from seeing their arms, hands, or the pistol.

The *S*s were given first a demonstration of the use of the shooting gallery and the method of setting the bar to appear level. They were allowed to shoot briefly and thus became aware of the operation of the buzzer and time clock. At this time they were told how to hold the pistol and urged to do as well as they could when shooting. The *S*s in the two supplementary groups were told that they would be unable to tell when they were on target since the buzzer would be disconnected and a "soundproof cover" placed over the relay and time clock to mask the clicking of these instruments. In actuality, these instruments were disconnected to insure that *S*s would receive no indication when they were on target.

Initial measurements of straight with the plain glass goggles and then with the prism goggles were made. Each measurement consisted of six settings made by *S* from alternate displacements of the bar by *E* to positions approximately 8 cm. above and below an apparently level position.

The average of the initial setting made with the plain glass goggles indicated *S*'s preexperimental perception of level. After the initial settings with the prism goggles were made, *E* either set the bar to the average of the initial measurements made with the plain glass goggles for *S*s in the apparently curved viewing conditions or to the average of the measurements made with the prism goggles for *S*s in the apparently straight viewing conditions.

All *S*s then viewed the target moving back and forth across the track for a period of 8 min. to allow time for the Gibson effect to develop for *S*s viewing the apparently curved bar. Following this viewing period all *S*s made another series of settings. It was assumed that 8 min. was long enough to achieve complete adaptation due to "normalization" and that subsequent changes in the settings in the apparently curved viewing conditions would reflect adaptation resulting from a change in efference.

There followed five 8-min. shooting periods separated by rest periods of 5 min. Immediately after each shooting period *S*s made a series of settings. A final setting with the plain glass goggles followed shortly after the last settings with prisms.

At the end of the experiment, which lasted for approximately 2 hr., *S*s were questioned about their impression of the goggles they wore, and the method they used to set the bar to appear level. All were paid $3.00 for their time.

Results

The six experimental groups were approximately equal on the initial measurements. The average magnitude of the measured prism-induced curvature is about 20.5 cm. for all of them. The measurements made after the initial nonshooting viewing period may be expected to reflect the Gibson effect for the apparently curved conditions. Those *S*s who viewed the target moving back and forth along an apparently curved path changed an average of 1.42 cm. in the direction of adaptation. The change for *S*s in the apparently straight viewing conditions was −.10 cm. The difference between the two viewing conditions was highly significant, $t(34) = 5.39$, $p < .001$.

The data used to test the major hypotheses were the differences between settings of apparently straight following the shooting periods and the settings which followed the initial nonshooting viewing period. This computation presumably removed the Gibson effect from the comparison between the apparently straight and apparently curved experimental conditions. The adaptation data for each shooting period are shown in Table 4. A negative sign indicates a change opposite to the adaptive direction, i.e., more perceived curvature.

Examination of the data in this table shows that the average amount of adaptation appears to vary nonsystematically from period to period. An analysis of variance of the increase in adaptation from the first two to the last two shooting periods produced no significant differences. The analysis of the data is, hence, presented using the most reliable single measure reflecting the effects of the experimental manipulations, namely, the average amount of adaptation for all five periods. The mean adaptation for the two infrared conditions was .27 cm.; for the two visible light conditions it was −.38 cm. These means were significantly different, $F(1, 32) = 14.19$, $p < .001$; and both were significantly different from zero by *t* test. There was no difference in the average magnitude of adaptation between the two apparently straight and the two apparently curved conditions. For the two apparently straight groups combined the adaptation was .02 cm.; for the two apparently curved groups combined it was −.14 cm. Apparently the initial nonshooting viewing period did eliminate the Gibson effect.

TABLE 4

MEAN ADAPTATION AFTER EACH SHOOTING PERIOD (IN CENTIMETERS)

Period	Experimental Cond.					
	Apparently Curved		Apparently Straight		Supplementary Groups	
	Infrared	Visible Light	Infrared	Visible Light	Aim Only	No Information
1	.32	−.59	.17	−.32	.36	−.20
2	.09	−.36	.25	−.07	.07	−.19
3	.03	−.59	.43	−.14	.05	−.62
4	.20	−.48	.49	−.45	.14	−.10
5	.28	−.29	.40	−.57	.07	−.07
Avg.	.19	−.46	.35	−.31	.14	−.24

The *S*s shooting with the visible light were on target an average of 49% of the total shooting time. Those shooting with the infrared light were on target only 18% of the time. This difference reflects the difficulty of hitting the target with the infrared light. The major question of interest concerning the performance data is the relative increase for the two shooting conditions. It was expected that *S*s shooting with the infrared light would improve over periods as they learned the correct arm and hand movements. Those shooting with the visible light were expected to improve very little. Their task was one which could be mastered rapidly. An index of relative improvement, the difference between the average hit time for the last two periods and the average for the first two periods divided by the sum of these two averages, was computed for each *S*. The difference between indexes for the shooting conditions was significant, $F\ (1, 32) = 11.55$, $p < .01$. The *S*s shooting with the infrared light showed more relative improvement. There was no significant difference between those who viewed apparently straight or apparently curved lines. There was, however, a significant interaction, $F\ (1, 32) = 5.39$, $p < .05$. In the apparently straight conditions there was a large difference between the shooting conditions; in the apparently curved conditions there was little difference.

It was expected that there would be significant positive correlations between the relative increase in performance and the amount of visual adaptation for *S*s in the infrared conditions. As these *S*s learned to issue appropriate efference they would be expected to both improve in performance and to visually adapt. The correlations between these two measures in the visible light conditions were expected to be negligible; improvement in performance was not expected to be associated with visual change since the efference issued in this condition was not associated with the perception of contour. None of these correlations, however, approaches statistical significance for any of the four groups.

The difference between the final and initial settings of straight with the plain glass goggles indicates the extent to which any visual adaptation persisted for "naked eye" measurements. As would be expected from the Gibson normalization phenomenon, there was a significantly larger aftereffect, $F\ (1, 32) = 5.55$, $p < .05$, for *S*s who viewed an apparently curved line (.65 cm.) than for *S*s who viewed an apparently straight line (−.02 cm.). The aftereffect data, unlike the adaptation data, include the normalization effect since they were computed from the initial settings of straight with the plain glass goggles that were made before the nonshooting viewing period. The difference between the shooting conditions, although in the expected direction, was not significant. The average aftereffect for *S*s in the infrared condition is .50 cm.; for *S*s in the visible light condition it was .13 cm. There was no interaction of the shooting and viewing conditions.

As was mentioned, two additional, apparently straight groups were run to clarify the findings and interpretations of the basic experiment. The Aim-only group was designed to test the conditions necessary for visual adaptation; the No-information group was designed to determine the changes resulting from prolonged viewing of the target and bar. The results of these two groups were analyzed by a one-way analysis of variance in conjunction with the results from the infrared and visible light apparently straight groups.

The average visual adaptation for each of these groups is also presented in Table 4. Using the average adaptation for all five periods, the four apparently straight groups differ significantly, $F\ (3, 32) = 3.54, p < .05$. The infrared group is significantly different from both the visible light group, $t(32) = 2.81, p < .01$, and the No-information group, $t(32) = 2.51, p < .02$. None of the other internal comparisons is significant. There is no significant difference in the aftereffect data between these four groups.

Discussion

The results continued to support a theory emphasizing the role of efferent readiness in determining the perception of contour. Those *S*s who had to learn to issue a new set of efferent responses to the perceived contour of the target's movement adapted to the curvature transformation significantly more than those *S*s who made approximately the same motor movements and had the same visual input but responded only to the discrepancy between the position of a visible spot of light and the target. Both the rate and path of the arm and hand movements were similar in these two conditions, but the responses of the *S*s shooting with the visible light more closely approximated the actual contour. These *S*s were on target almost three times as long as those shooting with the infrared light. The proprioceptive input from the hand and arm is, hence, clearly not the basis for visual adaptation.

The results further demonstrate that not all active or self-produced movement results in adaptation to curvature as Held's (1961) theory suggests. Instead, these results support the hypothesis that the important variable is whether or not the active movements are learned, so that the efferent readiness will be activated by a pattern of retinal stimulation. Atypical visual reafference is assumed to have occurred in both the visible light and infrared conditions, yet visual adaptation was obtained in only one condition. It can be argued that the necessity to aim, and the consequent attention paid to the position of the arm and hand in the infrared condition, resulted in more salient or usable atypical visual reafference than that which occurred from merely seeing the arm and hand in the visible light condition. The Aim-only condition was designed to clarify the distinction between adaptation resulting from a change in efferent readiness and adaptation resulting from a change in the correlation between self-produced movement and visual reafference. Since there was no buzzer to guide the arm movements made by *S*s in the Aim-only condition, more attention to the position of the hand would be expected in this condition than in the infrared condition with the buzzer feedback. Therefore, the Aim-only condition might result in maximum adaptation if this attention factor were critical. This group, however, showed no more visual adaptation than the apparently straight, infrared group, indicating that special attention to the hand and arm was not critical.

The negative adaptation found in the visible light condition was unexpected. The No-information group was run to test one obvious explanation. It was possible that continued viewing through prism spectacles resulted, in this situation, in increased perception of curvature. The average change for *S*s in the No-information condition is $-.24$ cm.; for *S*s in the visible light apparently straight group it is $-.31$ cm. These figures are very close and it appears that this shift does occur simply as a consequence of continued viewing in this situation.

Two of the findings were not in accord with the theoretical expectations, namely, the lack of significant differences between the shooting conditions in aftereffect and the lack of correlation between visual adaptation and performance improvement in the infrared conditions. Rapid decay of unstable adaptation may account for the lack of significant differences in aftereffect between the two shooting conditions, since about 5 min. elapsed between the end of the final shooting period and the beginning of the aftereffect measurements. The lack of expected correlation may be explained by the relative un-reliability of the measurements and the fact that the correlations are each based on only nine *S*s.

EXPERIMENT IV

If the efferent readiness that is activated by visual afferent input is important in determining the visual perception of contour, one might well expect that efferent readiness with respect to the extraocular muscles would be of particular importance. Considering the invariant relation that exists between eye movement and movement of stimulation across the retina, and considering the vast amount of experience that an individual has in establishing this relationship between input and output, it would not be surprising to find that efferent readiness relevant to eye movements was more intimately involved in visual perception than, for example, efferent readiness relevant to arm movements.

If this reasoning is correct, we would obtain, as we have obtained, only small amounts of change in visual perception of curvature when *S*s wear prism spectacles. Such spectacles produce a complex situation in which there is inconsistency between eye movements and other body movements that are evoked by the contour. If *S* engages in normal activity while wearing such spectacles, head movements, arm movements, and other body movements relevant to contour must conform to the objective shape. Thus, to the extent that these movements are in response to retinal input, *S* must learn new efference to associate with the visual input. He learns that he must move his head or his arm in a curved path in response to a straight pattern of retinal stimulation. These new afferent-efferent associations would presumably account for the observed change in the visual perception of curvature in the preceding experiments. There is, however, one major hindrance to change of visual perception. During the entire experience with the prism spectacles, the relationship between retinal input and efferent output to the extraocular muscles remains unchanged. The eyes, in order to achieve or maintain fixation, must move in conformity to the retinal image and *not* to the objective contour. Hence, to the extent that efferent readiness relevant to eye movements is important, this would interfere with and retard any change of visual perception when prisms are worn in spectacles.

If a situation could be arranged in which the movements of the eye had to conform to the objective contour rather than the retinal contour while wearing prisms, change of visual perception might occur much more quickly and dramatically. This situation can, indeed, be achieved by putting the prism on a contact lens rather than in a spectacle frame, as was realized by Taylor (1962) who arranged to have a scleral contact lens manufactured for his own right eye with a prism on it. He reported that, after he found the proper procedure for scanning contours to make adaptation rate maximal, his adaptation to curvature distortion was complete after only a short period of scanning.

There are reasons for not placing complete reliance upon this report. Taylor reported no data concerning the amount of curvature distortion produced by the prism on his contact lens other than to say that "... the distortion was less than I had hoped for [p. 227]." However, it is likely that the curvature distortion produced by Taylor's contact lens was very small. Because a prism on a contact lens is curved to conform to the curvature of the cornea, there is much less curvature distortion obtained than from a prism with a plane surface. For example, in the experimental work presented here 30-diopter prisms were used on contact lenses. The amount of curvature distortion produced was about comparable to what one would obtain from a prism with a plane surface of 4–8 diopters. Taylor's prism of not quite 12 diopters probably produced very little curvature distortion.

Nonetheless, the theoretical issue raised by Taylor appears important. If the efferent readiness relevant to eye movements is especially important, we should be able to find large and rapid changes in visual perception from wearing a prism on a contact lens even if the only movement in which the person engages are eye movements. It was, consequently, decided to replicate Taylor's study under more controlled conditions, using several *S*s who were completely naive as to what was happening and using prisms of large enough power to be sure that the curvature distortion would be clear and unmistakable.

Method

Subjects.—Three Stanford University students, two male and one female, were paid to serve in the experiment. They were told that the study would involve wearing a scleral contact lens for which they would have to be individually fitted. None realized that the contact lens produced curvature distortion.

Apparatus.—The lenses were manufactured by the Parsons Optical Laboratories of San Francisco, California. They succeeded in producing 30-diopter prisms on the contact lenses. The surfaces of the lenses were, of course, smoothed and rounded; none of the *S*s complained of any pain, none of them had any difficulty blinking or closing their eyes during rest periods. The lens and prism were cast in one piece out of optical plastic and then ground. Each *S* wore the prism in the right eye, which was also the dominant eye, with the base of the prism down. None of the *S*s had completely clear, sharp vision through the prism. There was some slight blurring.

The manufacture of the prisms was not easy and did not proceed without mishap. For the first *S*, it was thought that the contact lens would be stable with the prism base oriented laterally. It was manufactured in this way but, when the lens was first inserted into the eye, it immediately rotated so that the base of the prism was down. It was, however, stable in this position and that is how *S* wore it in the experiment. A serious error was made in the manufacture of the lens for the second *S* so that it did not fit at all. Fortunately, it was discovered that the lens manufactured for the first *S* fit the second one perfectly, also base downward, and so the first two *S*s actually used the same lens. No problems were encountered in manufacturing or fitting the lens for the third *S*.

The experimental apparatus was the same one used in the first two experiments, except that it was positioned on its side so that the rods were horizontal rather than vertical. This was done because the prisms, mounted base down, produced curvature of horizontal straight lines.

Procedure.—For the experimental sessions *S*s were seated in front of the apparatus with the head in a biteboard so that the right eye was directly in front of the center of the lines. When in the experimental situation, *S* saw only the two brass rods, the white background, and a small portion of the side of the frame of the apparatus.

Several experimental sessions were conducted with each *S*, each session lasting for approximately 90 min. At the beginning of each session *S*, with the left eye occluded and head on the biteboard, was asked to turn the knob on the apparatus so as to see the horizontal lines straight with his naked right eye. Four such settings were taken. On two of these measurements *E* displaced the line upwards from apparently straight before asking *S* to make the setting. On the other two measurements the lines were displaced downward from apparently straight. For a few of the last sessions with the second *S* and for all of the sessions with the third *S*, such initial measurements were also taken for the naked left eye with the right eye occluded.

The *S* then inserted the contact lens into his right eye and, again with his head in the biteboard, was asked to make four settings of the lines using the same procedure. Care was taken by *E*, moving the lines back and forth before *S* opened his eyes, to prevent *S* realizing that there was any curvature distortion. The subsequent procedure differed somewhat from *S* to *S*. The following aspects of the procedure were common to all three *S*s.

After these measurements, the lines were set by *E*, while *S*'s eyes were closed, either so that they were objectively straight, corresponding to the average of the settings *S* had made with his naked right eye, or apparently straight, corresponding to the average of the settings *S* had made with the contact lens in the eye. In each session *S* was then asked to simply look back and forth along the line. He did this usually for 5 min., was then once more asked to set the line so that it was straight, and was then given a 2-min. rest period during which he closed his eye and removed his head from the biteboard. The *E* reset the lines to the same position before the next period of looking back and forth along the lines. After a number of such periods, usually 8–10, the contact lens was removed and final measurements using the naked eye were taken.

The first few sessions for each *S* were conducted with the lines set so that they looked straight to *S*. Later sessions for each *S* were conducted with the line set objectively straight, so that they looked curved. The final session for each *S* was conducted with the line objectively straight, but with *S* looking monocularly through a prism in a spectacle rather than wearing the contact lens. This was done to get some information as to the magnitude of adaptation one might expect simply from the Gibson effect (Gibson, 1933) under these circumstances. The details of the procedure for each *S* follow.

*S*1

Session I. Lines set apparently straight. Ten periods of scanning the lines, periods ranging from 2 to 5 min. in length.

Session II. Lines set apparently straight. Ten periods of scanning the lines, periods ranging from 3 to 6 min. in length.

Session III. Lines set apparently straight. Five periods of scanning, each period 5 or 6 min. long. This was followed by two periods, one 3 min. and the other 6 min. long during which *E* moved a pointer along the line and *S* was asked to track the motion with his eye.

Session IV. Lines set apparently straight. Five periods ranging from 5 to 8½ min. in which *S* moved a stylus back and forth along the lines while looking. This was followed by two periods, each 6 min. long, in which *S* simply scanned back and forth.

Session V. Lines set objectively straight. Ten periods, 5 min. each, of scanning the lines.

Session VI. Lines set objectively straight. Six periods, ranging from 5 to 8 min., of scanning the lines.

Section VII. Lines set objectively straight. The *S* looking monocularly through prism in spectacle, 8 periods of 5 min. each of scanning the lines.

The procedure was more standardized for *S*s 2 and 3. All sessions contained eight periods of 5 min. each. The following tabulation gives the exact schedule for each of them:

	Sessions for *S*2	Sessions for *S*3
Lines apparently straight, scanning	I, II, & III	I, II, & III
Lines apparently straight, track pointer	—	IV
Lines objectively straight, scanning	IV & V	V, VI, & VII
Lines objectively straight, track pointer	VI	VIII & IX
Lines objectively straight, scan with prism in spectacle	VII	X

Results and Discussion

In spite of the occasional procedural differences among the three *S*s, it seems most sensible to present the data for all three together. In this way the uniformities among the three will most easily be seen.

Course of adaptation within a day.—Table 5 presents the data for the average daily adaptation to the prismatic distortion for the first three sessions when *S*s scanned an apparently straight line. There were no appreciable differences within any *S* in the course of adaptation among the different days and so the data are presented in terms of 3-day averages. The time point labeled "0" refers to the average settings made at the beginning of the session immediately after the contact lens was inserted in the eye. The time point labeled "10 min." refers to the averages of the setting made after the first 5 min. and after the second 5 min. of scanning the line. The row labeled "20 min." presents the average settings made after the third and the fourth periods of 5 min. of scanning, and so on. For *S* 1 there were minor deviations from this because his scanning periods were not always 5 min. long.

The last row of the table presents the percentage of adaptation calculated as the percentage of the distance between objectively straight and apparently straight that *S* adapted during the day. Thus, *S* 1 set the line as straight with the naked eye at 10.10; apparently straight at the beginning of the sessions with prism averaged to 12.00; the adaptation of .41 cm. divided by the initially perceived curvature of 1.90 cm. yielded the percentage of adaptation of 21.6. It is clear that all three *S*s showed adaptation during the course of the day. For the second *S* the total adaptation was already present after 10 min. while the visual change seemed more gradual for the other two.

It has been well known since Gibson's (1933) article that some perceptual "adaptation" occurs simply by looking at a curved line. If one studies adaptation to prismatically induced curvature by exposing *S* to apparently curved lines, one cannot easily separate this Gibson effect from other possible adaptation processes. It is, hence, not inconsequential to show that, with all three *S*s, definite and appreciable adaptation was observed to an apparently straight line—a situation in which the Gibson effect would not be contributing.

TABLE 5

Course of Daily Adaptation to Prismatic Curvature Distortion While Viewing an Apparently Straight Line

Time of Setting with Prism	Subject		
	1	2	3
0 min.	12.00	12.17	10.88
10 min.	11.80	11.95	10.86
20 min.	11.65	11.94	10.71
30 min.	11.59	11.97	10.71
40 min.	—	11.95	10.69
Naked Eye Setting at Start of Session	10.10	10.10	9.99
Percentage of Adaptation at End of Session	21.6	10.6	21.3

Note.—Average readings (in centimeters) are of settings of apparently straight. Three-day averages for each *S* are presented.

Each *S* also spent several sessions, one each day, exposed to an actually straight, and hence apparently curved, line. Each day the line was set according to *S*'s own initial naked eye measurements at the beginning of that session. Table 6 presents the data for the course of daily adaptation for each *S*, averaged for all the days in which *S* scanned the apparently curved line. It is clear from these data that all three *S*s showed large amounts of adaptation to the apparently curved line, over 40% after 40 min. of scanning. The rapidity and magnitude of these visual changes lend support to the idea that efferent signals to the extraocular muscles are heavily involved in determining perception.

In comparing the data in Tables 5 and 6, it is very clear that the adaptation to the apparently curved line was much larger than to the apparently straight line. In order to have a better basis for evaluating the magnitude of the adaptation effects obtained by scanning the apparently curved line, a measurement was made for each *S* of the Gibson effect on the last day on which *S* was run. The *S*s 1 and 2 wore 25-diopter prism spectacles while *S* 3 wore a 10-diopter prism spectacle. Each *S* was seated farther from the apparatus in an attempt to roughly match the measurement of curvature obtained from

TABLE 6

DAILY ADAPTATION TO PRISMATIC CURVATURE DISTORTION WHILE VIEWING AN APPARENTLY CURVED LINE

Time of Measurement with Prism	Subject		
	1[a]	2[a]	3[b]
0 min.	12.26	12.13	11.19
10 min.	11.73	11.44	10.79
20 min.	11.50	11.48	10.66
30 min.	11.24	11.38	10.55
40 min.	11.13	11.26	10.48
Naked Eye Setting at Start of Session	9.82	9.98	9.67
Percentage of Adaptation at End of Session	46.3	40.5	46.7

Note.—Averages of "apparently straight" settings on centimeter scale.

[a] 2 days.

[b] 3 days.

TABLE 7

ADAPTATION TO PRISMATIC CURVATURE DISTORTION WHILE VIEWING AN APPARENTLY CURVED LINE WEARING PRISM SPECTACLES

Time of Measurement with Prism	Subject		
	1	2	3
0 min.	12.52	12.78	10.84
10 min.	12.41	12.81	10.56
20 min.	12.26	12.40	10.46
30 min.	12.16	12.42	10.51
40 min.	12.18	12.31	10.39
Naked Eye Setting at Start of Session	10.05	9.93	9.68
Percentage of Adaptation at End of Session	13.8	16.5	38.8

Note.—Averages of "apparently straight" settings on centimeter scale.

the contact lens. The line was set in accordance with the initial naked eye setting of straight on that day and *S* spent the rest of the session scanning the line in 5-min. periods with his head fixed by a biteboard. The line was viewed monocularly with the same eye as was used with the contact lens. Table 7 presents the data on adaptation attributable to the Gibson effect and Fig. 1 presents a comparison of the data from Tables 6 and 7, that is, a comparison of adaptation with spectacles attributable to the Gibson effect and adaptation with the prism on a contact lens. It is quite clear that the absolute magnitude of the Gibson effect was about the same for all three *S*s. The large percentage figure for *S* 3 was due to the small magnitude of the initial curvature that was produced. The amount of adaptation with the contact lens was greater for all *S*s than the magnitude that could be attributed to the Gibson effect.

Effect of saccadic and smooth tracking eye movements.—There is some evidence that would lead us to expect that, if the efferent readiness activated by a given retinal input determines the visual perception of contour, then adaptation to curvature distortion with a prism on a contact lens would be greater if the eye engages in saccadic movements than if the eye follows the contour with a smooth tracking movement. Rashbass (1961) reported that these two types of eye movements were controlled by different mechanisms in the central nervous system and that, while the saccadic movement was a fixation response, the smooth tracking movement was a response to direction and velocity of movement across the retina. This perhaps suggests that ef-

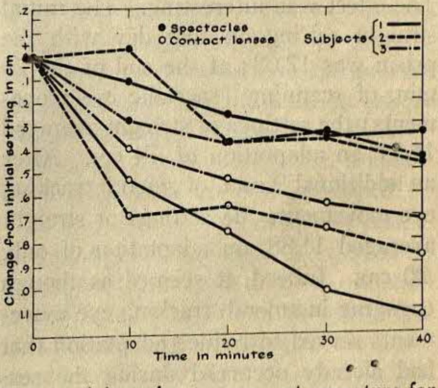


FIG. 1. Adaptation to apparent curvature for prisms in spectacles and on contact lenses.

ference issued to produce the smooth tracking motion is not coded in terms relevant to contour while the efference to produce a saccadic movement, or a series of them, would be more concerned with contour, that is, with specific position in space of the point fixated relative to the point of fixation from which the eye moved. Festinger and Canon (1965) reported data which lend some support to this. They showed that after saccadic eye movements there is better localization of the fixated point in space than after a smooth tracking eye movement.

Some exploration was done to investigate this suggestion. During sessions in which *S* was to engage in smooth tracking eye movements *E* moved a pointer back and forth along the line and asked *S* to follow the pointer. The movement of the pointer in this way obviously did not provide a very uniform smooth motion for *S* to track and, over a 5-min. period there was, undoubtedly, many saccadic movements, but it was felt that this procedure might, at least, provide *E*s with some preliminary information.

The third session for *S* 1 began as a free scanning session with an apparently straight line. After 29 min. of such scanning, *S* spent 9 min. following the tip of the pointer as *E* moved it. The effect was interesting. The initial straight setting on that day with the prism was 12.08; at the end of the 29 min. of scanning (saccadic eye movements) the settings of straight averaged 11.34, an adaptation of .74 cm. After an additional 9 min. of smooth tracking eye movements the settings of straight averaged 11.88, an adaptation of only .20 cm. Indeed, it seemed as though engaging in smooth tracking eye movements served to reduce adaptation that had already occurred during the session. No more "tracking" sessions were conducted with this *S*. This small bit of data was interesting enough, however, so that we explored it more systematically with the next two *S*s.

The *S* 2 spent one session tracking the pointer while viewing an apparently curved line. At the end of this session there was .67 cm. (30.6%) adaptation to the curvature. This is to be compared with .87 cm. (40.5%) adaptation on the previous 2 days when scanning the apparently curved line. The *S* 3 spent one session tracking while viewing an apparently straight line. There was no adaptation whatsoever during this session. The measurement at the end of the session showed a change of −.27 actually in the opposite direction from adaptation. He also spent two sessions tracking while viewing an apparently curved line. The average adaptation for these two sessions was .65 cm. (38.0%). The comparable figure for the previous 3 days of scanning was .71 cm. (46.7%). Thus, in all instances, there was less adaptation when tracking with smooth eye movements than when scanning with saccadic eye movements. The difference was smallest for *S* 3 when viewing the apparently curved line and greatest for this same *S* when viewing the apparently straight line.

Aftereffects of adaptation.—At the beginning, and at the end, of every session *S* was asked to set the line so that it was straight with the naked eye. By comparing these two sets of measurements, one can see whether or not there were aftereffects for the naked eye of the adaptation occurring during the session while *S* wore the contact lens. Table 8 presents these data. Examination of the data shows that, except for the first day, there was an aftereffect on each day for each *S*. After the first session, all three *S*s, on every day, showed a change from initial

TABLE 8

Aftereffects of Adaptation for the Naked Eye After Wearing Contact Lens

	Subject 1		Subject 2		Subject 3	
	Pre	Post	Pre	Post	Pre	Post
Day 1	10.02	10.11s	9.99	9.78s	9.94	9.96s
Day 2	10.22	10.14s	10.22	9.75s	10.09	9.78s
Day 3	10.06	9.78s	10.10	9.74s	9.94	9.76s
Day 4	9.99	9.80sv	10.04	9.41c	9.93	9.64st
Day 5	9.93	9.51c	9.92	9.27c	9.81	9.41c
Day 6	9.71	9.43c	9.89	9.26ct	9.69	9.33c
Day 7					9.51	9.27c
Day 8					9.54	9.26ct
Day 9					9.57	9.28ct

Note.—Averages of "apparently straight" settings on centimeter scale. s = viewed apparently straight line, c = viewed apparently curved line, t = tracked pointer, v = on this day *S* 1 moved a stylus along lines himself.

to final measurements with the naked eye in the same direction as the adaptation changes.

The data revealed another interesting result. There seems to be a cumulative effect over days in the settings made at the *beginning* of the session with the naked eye. Each of the three *S*s shows the same pattern. From the first to the second day the settings changed in a negative direction, that is, opposite to what would be expected on the basis of adaptation to the prism. From then on, however, each day showed a progressive effect, the naked eye settings at the beginning of the session becoming progressively more and more curved in the direction expected from adaptation.

There is no compelling reason to expect a carryover of adaptation to the prism from day to day. The *S* spent less than an hour each day looking through the prism and then had some 15 hr. of normal vision in which to readapt. It is, on the other hand, perfectly possible that a conditional adaptation and a conditional aftereffect might develop. That is, the stimulus conditions of the head in the biteboard in front of the apparatus could revive, or have become specifically associated with, the adaptation to the prism curvature and the progressive day to day cumulative effect could be due to the gradual development of this conditional learning.

This would be an attractive interpretation except for the fact that there exist data which make it implausible. If the progressive changes in initial settings with the naked eye at the beginning of each session are, indeed, due to conditional adaptation, then one would also expect to observe a similar cumulative effect on the first settings made on each day with the contact lens in the eye. No such trend for these settings was observed. Indeed, if anything, there was a tendency to perceive more and more curvature rather than less and less. It is difficult to believe that a conditional aftereffect had developed while, at the same time, a conditional adaptation had not developed.

Interocular transfer of aftereffects.—About midway through the period of experimentation with *S* 2, it occurred to *E*s that it would be interesting and valuable to obtain data on interocular transfer. Taylor (1962) implied that the aftereffects of adaptation with a prism on a contact lens did not transfer. He wrote: "The clearest evidence of adaptation came about half an hour after removal of the lens. Sitting in a stationary car I got the impression that the vertical lines of a tall building in front of me were not quite straight. I then inspected the lines monocularly, and found that with the right eye they were straight but with the left eye they were curved, the convexity being to the left, that is, the opposite of what I had seen through the lens [p. 225]."

Theoretically, the issue is important. If the change in visual perception is actually due to a change in the efferent readiness activated by the retinal input, then there is every reason to believe there would be considerable, if not complete, interocular transfer.

The first time that initial and final measurements with each naked eye were taken was in Session V for *S* 2. The change from initial to final setting for the naked right eye (the one that wore the contact lens during the session) was +.65 cm.; for the left eye the change was +.33 cm., about 50% transfer of the aftereffect from one eye to the other. These measurements were repeated in Session VI in which *S* 2 visually tracked the pointer that *E* moved. This time the aftereffect for the right eye was .63 cm. but for the left eye was only .07 cm., indicating virtually no transfer at all.

Interocular transfer of the aftereffect was measured systematically for every session with *S* 3. With this *S* there is no question but that the aftereffect transferred completely from the right to the left eye. The simplest way to present the data is by showing, on the same figure, the initial and final naked eye measurements on each day for each eye. This is shown in Fig. 2. It is clear that there was about 100% interocular transfer on every day except Day 8 when the transfer was not quite as large, only about 75%.

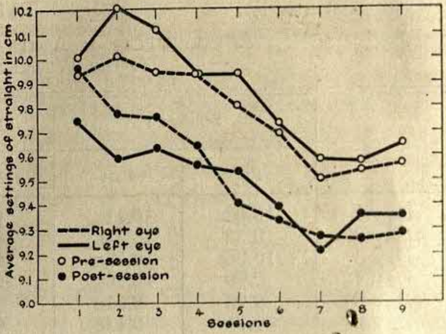


FIG. 2. Aftereffect and transfer of adaptation to apparent curvature for *S* 3.

DISCUSSION AND CONCLUSIONS

Toward the end of the nineteenth century many psychologists held that the efferent system was importantly related to consciousness. Such theories fell into disrepute, however, and remained only as historical curiosities.

Recent evidence, both of a psychological and physiological nature, indicates the value of reexamining such theoretical positions. Several persons have also, recently, proposed new theories that afford efferent activity and readiness for efferent activity an important place in determining conscious perception.

We have reviewed the evidence relating to this question and have stated an experimentally testable theory of visual perception of contour. This theory, which seems to fit known facts, holds that visual perception of contour is determined by the particular sets of preprogrammed efferent instructions that are activated by the visual input into a state of readiness for immediate use.

Four experiments were done to test whether or not the conscious experience of visual perception is determined by the efferent readiness activated by the visual input. In three of these experiments *S*s wore prism spectacles producing apparent curvature of straight lines and made arm movements corresponding to the objective contour of the lines while viewing them through the prisms. In each experiment one set of experimental conditions was designed to facilitate learning to issue new efference to the arm in response to the retinal contour and one set of conditions was designed to hinder such learning of a new afferent-efferent association. In all three experiments there was significantly more change in the visual perception of "straight" in the conditions that encouraged learning a new afferent-efferent association.

Theoretically the data support the view that visual input activates a whole set of learned efferent readinesses and that these latter determine the conscious experience of visual perception. In these three experiments *S*s in the Learning conditions learned to issue efference to the arm appropriate to a curved path in response to visual input of a straight path. During the measurements of the perception of straight no relevant arm movements were involved. Consequently, it seems reasonable to conclude that, having learned such a new afferent-efferent association, the visual input of a straight line now activates efferent readiness relevant to arm movement corresponding to a curved path. There is a consequent change in the visual perception.

In a fourth experiment change in the visual perception of curvature was measured for three *S*s who viewed a line monocularly through a wedge prism mounted on a contact lens. For each *S* the head was fixed by a biteboard and the only movement relevant to the contour was movement of the eyes. There is an important difference in the pattern of eye movements that *S* must make depending on whether the prism is mounted in spectacle frames or on a contact lens. In the former case, the eye in scanning the contour must move in accordance with the retinal image if fixation is to be maintained along the contour. However, if the prism is mounted on a contact lens so that the face of the prism moves as the eye moves, then the eye movements must conform to the objective contour in order to maintain fixation along the line. Under these circumstances the old, well-learned efference for an eye movement to fixate that is activated by visual input will result in a loss of fixation. To move the eye and maintain fixation along the contour, *S* wearing the contact lens must learn a new set of efferent instructions to issue in response to the visual input. If the conscious experience of visual perception of contour is, indeed, determined by efferent readiness activated by the visual input, then to the extent that *S* learns a new afferent-efferent association and, hence, a different efferent readiness is activated by the visual input, he will have a different visual perception of the contour.

In accordance with these theoretical expectations all three *S*s showed appreciable change in the visual perception of curvature as a consequence of simply scanning the line while wearing the contact lens. This occurred whether *S* viewed an apparently straight line or an apparently curved line. Further evidence suggests that there is appreciable, perhaps complete, interocular transfer of this change in perception of contour. The data also provide a hint that, if the eye movement involved is a smooth tracking movement, there is less change in visual perception than if the eye movements are saccadic.

While the data are not conclusive with regard to an "efference readiness" theory of visual perception, they do support the theory. All four experiments taken together provide considerable evidence that such a theory has some validity and merits further consideration and exploration.

REFERENCES

Breese, B. B. On Inhibition. *Psychol. Rev. Monogr. Suppl.*, 1899, **3** (1, Whole No. 11), 1–65.

Campbell, F. S., & Robson, J. G. A fresh approach to stabilized retinal images. In, *Proc. Physiol. Soc.*, 1961, 11P–12P.

Campbell, D., Sanderson, R. E., & Laverty, S. G. Characteristics of a conditioned response in human subjects during extinction trials following a simple traumatic conditioning trial. *J. abnorm. soc. Psychol.*, 1964, **68**, 627–639.

COHEN, M. Visual curvature and feedback factors in the production of prismatically induced curved-line after-effects. Paper presented at the meeting of the Eastern Psychological Association, New York, April 1963.

COHEN, W. Form recognition, spatial orientation, perception of movement in the uniform visual field. In A. Marrin & E. P. Horne (Eds.), *Visual search techniques.* (Publication 712) Washington: National Academy of Science-National Research Council, 1960. Pp. 119–123.

FENDER, D. H. The eye movement control system: Evolution of a model. In R. R. Reiss (Ed.), *Neural theory and modeling.* Stanford, Calif.: Stanford Univer. Press, 1964. Pp. 306–324.

FESTINGER, L., & CANON, L. K. Information about spatial location based on knowledge about efference. *Psychol. Rev.,* 1965, **72,** 373–384.

GIBSON, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.,* 1933, **16,** 1–31.

HARRIS, C. S. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychol. Rev.,* 1965, **72,** 419–444.

HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.

HELD, R. Exposure-history as a factor in maintaining stability of perception and coordination. *J. nerv. ment. Dis.,* 1961, **132,** 26–32.

HELD, R., & REKOSH, J. Motor-sensory feedback and the geometry of visual space. *Science,* 1963, **141,** 722–723.

HELMHOLTZ, H. VON. *Physiological optics* (Trans. by J. P. C. Southall) New York: Dover, 1962.

JAMES, W. *The principles of psychology.* Vol. 2, London: Macmillan, 1890.

KOHLER, I. The formation and transformation of the perceptual world. (Trans. by H. Fiss) *Psychol. Iss.,* 1964, **3**(4), 175.

MERTON, P. A. Human position sense and sense of effort. Society of Experimental Biology Symposium XVIII, *Homeostasis and Feedback Mechanisms.* Cambridge: Cambridge Univer. Press. 1964, Pp. 387–400.

MONTAGUE, W. P. Consciousness: A form of energy. *Essays, philosophical and psychological, in honor of William James.* New York: Longmans, Green, 1908. Pp. 103–105.

MÜNSTERBERG, H. The physiological basis of mental life. *Science,* 1899, **9,** 442–447.

MÜNSTERBERG, H. *Grundzuge der Psychologie.* Vol. I., Leipzig: J. A. Barth, 1900.

PICK, H., & HAY, J. Adaptation to prismatic distortion. *Psychon. Sci.,* 1964, **1,** 199–200.

RASHBASS, C. The relationship between saccadic and smooth tracking eye movements. *J. Physiol.,* 1961, **159,** 326–338.

SHERRINGTON, C. S. The muscular sense. In E. A. Schafer (Ed.), *A textbook of physiology.* Edinburgh & London: Pentland, 1900.

SPERRY, R. Neurology and the mind-brain problem. *Amer. Scient.,* 1952, **40,** 291–312.

TAYLOR, J. G. *The behavioral basis of perception.* New Haven: Yale Univer. Press, 1962.

TEPAS, D. I. The electrophysiological correlates of vision in a uniform field. In M. A. Whitcomb (Ed.), *Visual problems of the armed forces.* Washington: National Academy of Science-National Research Council, 1962. Pp. 21–25.

VON HOLST, E. Relations between the central nervous system and the peripheral organs. *Brit. J. Anim. Behav.,* 1954, **2,** 89–94.

WASHBURN, M. F. *Movement and mental imagery.* Boston: Houghton Mifflin, 1916.

WUNDT, W. Zur Theorie der Raumlichen Gesichtswarnehmungen. *Philos. Stud.,* 1898, **14,** 11.

(Received June 6, 1966)